FACHHOCHSCHULE KÄRNTEN

Master Thesis

Spatio-Temporal Data Mining for Pattern Recognition in Production Line Processes

by

Stefan Schabus, BSc

Submitted in partial fulfilment of the requirements of the degree
Master of Science in Engineering

Spatial Informaton Management
Carinthia University of Applied Sciences
School of Engineering & IT
Department of Geoinformation & Environmental Technologies

Internal Supervisor:

Dr.-Ing. Karl-Heinrich Anders, School of Spatial Information Management, Carinthia University of Applied Sciences, Carinthia, Austria

Mag. Dr. Gernot Paulus MSc. MAS, School of Spatial Information Management, Carinthia University of Applied Sciences, Carinthia, Austria

External Supervisor:

Prof. Dr. Andrè Skupin, Department of Geography, San Diego State University, San Diego, United States of America

Dr. Johannes Scholz, iSpace, Researchstudio Austria, Salzburg, Austria

Villach, 22nd June, 2014

Statutory Declaration

I hereby declare that:

- the Master Thesis has been written by myself without any external unauthorized help and that it has not been submitted to any institution to achieve an academic grading.
- I have not used sources or means without citing them in the text; any through from others or literal quotations are clearly marked.
- the enclosed master Thesis is the same version as the version evaluated by the supervisors.
- one copy of the Master Thesis is deposited and made available in the CUAS library (§8 Austrian Copyright Law [UrhG]).

I fully understand that I am responsible for the application for a patent, trademark or ornamental design and that I have to prosecute any resultant claims myself.

Villach, 22nd June, 2014

Place, Date | Signature

Acknowledgements

First of all, I would like to thank my parents and my family for supporting me and providing me with perfect opportunities for my whole education and study. Further, all my friends deserve a certain credit for making the years of study unforgettably great and encouraging me to complete my study abroad.

I am deeply grateful and thankful to my supervisor at the San Diego State University, *Dr. Andrè Skupin* for giving me the chance to finish my Master studies in California. He gave me an excellent support throughout my stay at SDSU and made me aware of Self-Organizing Maps. He has a great ability to motivate and to encourage people for his work and showed many new and creative scientific researches, opening my eyes and my mind for new topics. It is also important to thank his research group, which was very helpful with many valuable discussions which improved my work and was also a big motivation. Thanks for all!

My appreciation refers also to my internal supervisors *Dr. Gernot Paulus* and *Dr.-Ing. Karl-Heinrich Ander*s from the Carinthia University of Applied Sciences. Their feedback and support gave me the basis for my research and lead me into the right direction.

I would also thank *Dr. Johannes Scholz* for his generous support throughout the whole research. I appreciate his way to motivate and the helpful and valuable discussions. Thank you very much for the support and the feedback!

An important role played the *Austrian Marshall Plan Foundation*. Without their financial support and covering a great share of my costs of the stay in San Diego, I would not have been able to come to San Diego and to gain new knowledge. I am deeply thankful for the funding.

Further, this thesis is supported by *Infineon Technologies Austria AG* by providing information about the production line. Their input data was the basis for the whole research and many knowledge was provided enabling the research. I am also very thankful, as they supported the data preparation and migration and some basic approaches financially.

Abstract

Transportation of production assets in complex production line processes is one critical issue regarding the productivity of a manufacturing site. Such transportation tasks in a flexible, highly dynamic production line environment are often performed by humans.

The major goal of this master thesis research project is to provide spatial decision support by analyzing the spatio-temporal patterns of production asset trajectories and quality measures for production line processes. In a first step the overall indoor geography of a production line process is conceptually modeled and the indoor environment under review is prepared for the developed model. A new created production line specific transportation network combined with already existing tracking data for production assets provide important base data sets for this research project. Based on a spatio-temporal data model that has to be developed all relevant data are captured. In a further step, the provided spatio-temporal data is analyzed in order to identify and to understand critical zones in the production line. Such critical zones can be areas slowing down the production process or areas having an impact on the quality of production assets. The conceptualization and definition of the transportation and quality dependent data is a crucial issue for the analysis. Spatial data mining methods like Self-Organizing Maps *(SOM)* will be applied on the base data to analyze and visualize transport and quality dependent spatio-temporal patterns and their change during the production process.

The results of this research will be critically evaluated by production domain experts. The expected results of this master thesis research project will be on the one hand the integration of all indoor geography and process relevant data like transportation network, production asset trajectories and quality parameters to a spatio-temporal data base. On the other hand is the analysis of quality and movement patterns based on the prepared spatio-temporal data such as production asset trajectories in the production line processes. A final evaluation of the spatial data mining results will provide new knowledge for a final recommendation about a general workflow and all necessary steps and resources in order to set up and implement the basis for an indoor geography for a production line, including production asset trajectories and quality issues, to improve the performance of a production line and to show new ways to monitor and proof quality issues of production assets.

Keywords: *Spatio-temporal Data Mining, Indoor Geography, Trajectory Analysis, Production Line Processes*

Contents

List of Abbreviations

BIM	....	*Building Information Model*
BMU	....	*Best Matching Unit*
CUAS	....	*Carinthia University of Applied Sciences*
FME	....	*Feature Manipulation Engine*
GIS	....	*Geographic Information System*
GNSS	....	*Global Navigation Satellite Systems*
GPS	....	*Global Positioning System*
RFID	....	*Radio Frequency Identification*
SOM	....	*Self-organizing Maps*

Chapter 1

Introduction

The fundamental issue of this Master Thesis is the investigation of an innovative way visualizing and analyzing movement behavior of production assets being transported between between several production line processes of a highly flexible production line. A prototype implementation is used to show an approach how spatio-temporal data, collected in a production line, can lead to new knowledge about movement behavior and quality issues. Therefore, a complex workflow leads to an advanced knowledge and expertise out of loads of data by visual analytics and generated maps. This Master Thesis comprises research topics how outdoor Geographic Information Systems *(GIS)* are applied in an indoor environment taking into account real-time tracking. Further a fundamental part of the research is the spatio-temporal data recorded in a production line including a spatio-temporal database and a spatio-temporal data mining for this specific field of a production line. Generally, this approach of spatio-temporal data mining and pattern recognition combines the topics of a production line, indoor geography and the data mining has the ability to open a new field of application for GIS.
This introduction chapter provides an overview of the main motivation for this research which combines the analysis and optimization of a production line and the comprehensive research topic of indoor geography. A brief delineation of the research problems and project goals is provided as well as a description of the proposed method of solution. The expected results are summarized followed by the audience of this research. Finally, the structure of this Master Thesis is described.

1.1 Motivation

A significant number of research projects were carried out over the last decades in the context of modeling outdoor space providing a rich set of methods high level of structuring and applications. Hence, in the last years more achievements have been made in outdoor geography than indoor geography. However, indoor geography related research has attained increasing attention during the last years. Thereby, various studies showed that an average person spends about 90 percent inside a building *(Worboys, 2011; Giudice et al., 2010)*.
The increasing size of buildings combined with an increasing complexity of build-

ings shows the need for indoor location based services, such as indoor maps or even indoor routing especially in public buildings *(Goetz, 2012)*. Modeling the indoor structure of buildings for spatial applications is quite new and not a straight forward task ,regarding to the complex, high resolution building structure, the functionality of the building, the accessibility and also the field of application for which the building must be structured *(Meijers et al., 2005; Ascraft, 2008)*.
In addition to that, production line processes represent another new challenging research topic and application field of indoor geography. Optimization of production processes depends on allocation and sequencing of production processes. Such optimization will potentially increase the efficiency of a production line and therefore provide an interesting option for cost savings and an increase of performance and productivity *(Nyström et al., 2006)*.
A huge amount of spatio-temporal data is gathered in companies storing historical information such as the movement behavior of production assets. Therefore, spatio-temporal data mining can have a fundamental issue on gaining new knowledge in large datasets. Maps can be created to visually analyze the results of the data mining by human. In this case, the main advantage is that the human has the ability to recognize visual patterns *(Compieta et al., 2007)*. The recognized patterns are useful for the optimization by identifying already optimal areas or to highlight areas with potential to increase the processes and make them less error prone. Visual analysis of quality parameters can lead to new and useful information to decrease the amount of errors or quality differences in products.
A further motivation for intelligent pattern recognition of movement behavior or quality issues in production line processes is the facility management. Proper maintained and implemented models can become a powerful tool, which can be used to manage and to operate the entire structure of a facility. Determining upgrades in the sequence of the processes in the production line can enhance the cost-effectiveness and is an advantage of such an implemented system *(Ascraft, 2008)*.

1.2 Research Problems and Project Goals

The indoor geography on its own is a very challenging topic, especially if it describes the indoor environment of a production line. Therefore, an own building structure has to be modeled comprising the indoor geography of a production line. Further is the tracking of the different transportation processes in a flexible, highly dynamic production line. Trajectories and the movement behavior are important to identify moving patterns and data clusters *(Bogorny et al., 2006)*.
Spatio-temporal data describing the movement behavior of production assets needs an own data structure and a spatio-temporal database for a proper analysis. This is challenging in terms of combining space and time in one central data model *(Fan et al., 2010; Lee et al., 2009)*.
The spatio-temporal pattern recognition of the movement behavior in the produc-

tion line is a new way, using intelligent visualization possibilities to gain new knowledge out of large datasets. This helps to figure out certain bottle necks and to investigate transportation stops in the trajectories through the production line (Alvares et al., 2009).
To sum up, the main goals are to set up an indoor environment for a production line and to set up a spatio-temporal data structure to enhance the spatio-temporal analysis of movement behavior and quality issues. This is including a spatio-temporal data structure and database to answer predefined questions. All this goals are corresponding to the main research problems.

1.3 Methods of Solution

The proposed method of solution starts with an exhaustive literature review to gain a deeper insight into the stated research problems. The literature review is mainly concerning the general positioning outdoor and indoor, knowledge about storing and handling spatio-temporal data and the capability of routing in an indoor environment such as a production line. This provides the basis for the development of the spatio-temporal data structure and the indoor environment as such. Another important part of the literature review concerns analysis of movement paths and trajectories providing insight in the data management of trajectory data and the final part is getting knowledge about Self-Organizing Maps which is one spatio-temporal data mining method.
The gained expertise is used in the further pre-analysis of the provided spatio-temporal data of a production line, the production line under review and the indoor positioning solution as well as a general description of the provided base data. This leads to the modeling phase, which is divided into the use case modeling, the development of a spatio-temporal data structure and a characterization of the indoor environment of a production line. The use case modeling provides an abstract description of the actors interacting in the production line environment and highlights questions that have to be answered through the further developed spatio-temporal data model and the indoor environment. This leads to the modeling of an indoor ontology very specifically describing the indoor environment of a production line in an abstract way. The spatio-temporal data model is resulting out of the pre-analysis of the existing data and the use case model.
To implement the prototype, the spatio-temporal database is developed which is the basis for the following data preparation and data migration of the provided input data. After that, a graph based network is set up to enable the routing in the indoor environment and to create trajectories of the transport ways through the production line. Hence, the data analysis is carried out which starts with an evaluation of existing data mining software and existing projects carried out by students of the Carinthia University of Applied Sciences *(CUAS*) one year ago. Additionally, the data modeling for the spatio-temporal data mining is carried out including the movement behavior and the quality issues. This is necessary to be able to interpret

the results of the analysis. Finally, spatio-temporal data mining is carried out analyzing the modeled movement behavior and quality issues as well as a comparison of the physical space and the attribute space.
The visualization is done using one type of visualization of separate tracks and only parts of transport ways to be able to compare the movement behavior of the production assets, giving a basic insight and understanding of the tracks. Further Self-Organizing Maps are created out of the spatio-temporal data mining and the results are visualized either as component planes as well as a website having the capability of comparing the physical space, the attribute space and the time. In the end the results are validated with expert knowledge and the approach is proofed.

1.4 Expected Results

The result of this Master Thesis will be a prototype implementation visualizing the movement behavior and patterns of production assets in a highly flexible production line. Further, it will also be possible to directly compare the physical space of the production line with the quality space of production assets moving through the production line over time.
To get the expected results, challenging intermediate results are as follows:

- Modeled use cases and an ontology to show how to set up an indoor geography for a production line.
- Spatio-temporal data base concept including all the modeled aspects for an indoor geography. Very useful for further historical questions such as where was an equipment last year.
- A network dataset combining all the production line processes as well as indoor and outdoor geography for any further tasks that can take place. It is necessary to route through the production line.
- PostGIS functions to snap the tracked positions of the goods to the network and enable the data mining of transport patterns. Quality patterns are also analyzed along the created network.
- Spatio-temporal data mining of the transport and moving behavior and the quality issues as self-organizing maps.
- Comparison of the physical and attribute space of quality parameters of a single devices and global in the production line.
- Map generation of the identified patterns in the production line.
- Validation of results by domain experts.

Therefore, the research paper deals with spatial data mining tasks on moving patterns. Trajectories will be analyzed, in order to identify bottlenecks or problems in the moving behavior, occurring in a production line to improve production assets. Further, data mining tasks belong to quality parameters of the production assets in the production line. The recognized transport patterns and quality patterns are visualized through self-organizing maps that can be helpful to plan improvements. This is an approach bringing outdoor GIS indoor with respect to real-time tracking and optimization of a production line. The approach will discuss and evaluate the overall potential of spatio-temporal modeling and data mining in order to increase effectiveness, quality assessment and cost efficiency in production line processes.

1.5 Audience

This research addresses primarily industries with highly flexible production lines, such as semiconductor power fabs. There, the transportation of production assets from one production step to the next production step is a crucial issue for the productivity. Most manufacturers try to automate the transport process in order to speed up the processes and make them less error prone. Due to the fact, that in a flexible, highly dynamic production line the transport processes are done by humans there is room for improvement of the transport processes, by providing (spatial) Decision Support in real-time.
Another interested group is the general research because outdoor GIS is used in an indoor environment. Further, the indoor geography is combined with research topics in spatio-temporal databases and spatio-temporal data mining tasks.

1.6 Structure of the Thesis

The second chapter states the hypothesis for the research project. This is including the scope of the thesis and also what is not scope of the thesis.
The third chapter gives a detailed description on the theoretical background of the project. These are the positioning outdoor and indoor, the handling of spatio-temporal data, routing in an indoor environment and related projects.
The methodology of the project is stated in the fourth chapter including a general workflow for the proposed solution, the indoor positioning solution and a description of the base data. The modeling phase is also part of this research approach whereas the use cases, the spatio temporal data concept and the indoor environment is modeled.
The fifth chapter focuses on the implementation of the prototype including the spatio-temporal database, a routing network and the creation of the transport ways. Data mining is included with mainly self-organizing maps of the movement behavior and quality assets of production goods.
Chapter 6 is outlining the results of the used approach with the indoor environment and the routable network. Second is the comparison and analysis of the data mining tasks with the self-organizing maps.
Chapter seven discusses the results and is a proof of the use concept. Chapter eight summarizes the approach and provides a conclusion as well as an outline for future perspectives.

Chapter 2

Hypothesis and Scope of the Thesis

The hypothesis of this thesis is bringing the outdoor geography in an indoor environment, which is a highly flexibly production line environment. This production line environment comprises spatio-temporal data which is recorded during the transportation process of production assets as well as quality issues. By the comprehensive task of bringing the outdoor geography into the indoor environment, it is possible to execute spatio-temporal data mining techniques to analyze the recorded data of the production line.
The scope of this research is in the context of indoor geography and production line processes is:

- Modeling of a building structure which is very specific in different fields of application and the complexity of the buildings itself. This is then combined with a visualizing problem of the applications which is most of the time carried out in 2D and not in a 3D visualization environment which is much more effective as buildings have also several floors.
- The localizing or tracking of the transportation processes in a flexible, highly dynamic production line. The trajectories of the production assets are directly linked with data mining approaches, which are important to identify relevant and important tracks and data clusters *(Bogorny et al., 2006)*. Tracking and positioning of production assets is not a task of this research in particular as a Radio Frequency Identification *(RFID)* based positioning system *Thiesse et al. (2006)* is already implemented and the data are provided.
- There exists a high variability in the production processes, in terms of number of production assets and production processes to be carried out. In order to work with a homogeneous base dataset, a fixed time frame is extracted to enable the analysis. This specific period of time can either be the whole process steps or process steps in between. So the process steps start before the selected time frame and can also end after the selected time frame.
- To enable the data mining or any further analysis a concept for a spatial-

temporal database has to be developed which is very challenging as space and time have to be included in one model *(Fan et al., 2010; Lee et al., 2009)*.

- Data Mining is challenging to figure out bottle necks, stops and candidate stops *(Alvares et al., 2009)*. The trajectories itself have to be linked along a network providing useful information of how the production line is linked.
- Evaluation of results is necessary to proof the developed concept and to validate the correctness of the spatio-temporal data mining. This is challenging in case of the existing large datasets that have to be compared with the actual results of the analysis.

To sum up, the scope includes the setting up of an indoor geography describing the highly flexible production line as indoor environment. Spatio-temporal data leads to the creation of different trajectories of production assets in a network as well as the modeling as well as quality data will be conceptually modeled and stored in an implemented spatio-temporal database. Based on the prepared and structured input data, spatio-temporal data mining is carried out to enable the analysis to identify for example bottle necks and moving patterns during the production line processes. Additionally, a comparison of the physical space, attribute space and the time will be analyzed by Self-Organizing Maps.
Not covered in this thesis is the set up of an indoor positioning system and no evaluation of the accuracy of the data. The base data for a production line and the positioning data is provided. Also not scope of this thesis is the shortest path algorithm, where an already existing algorithm is used.

Chapter 3

State of the Art and Relevant Literature

This chapter is reflecting the relevant literature and state of the art of common projects and techniques. First is the positioning covering outdoor positioning and indoor positioning, whereas there is an extra focus on indoor positioning in a production line. Second is the spatio-temporal data including the storage of spatio-temporal data and spatio-temporal data mining capabilities. Additionally, this section covers literature on routing in the indoor environment and ontologies for indoor environment purposes.

3.1 Positioning

The capability to identify the precise position or location of a person, a machine or anything else in real-time is important and crucial for a GIS as it is for example the basis for a guidance system *(Mautz, 2008)*. The importance of an exact and accurate position is increasing in various applications and society *(Barnes et al., 2003)*. Therefore, positioning is a comprehensive task by taking into account the indoor and the outdoor environment *(Li et al., 2008)*. One of the main problems is the different focus. Indoor positioning is focusing on limitations in size of rooms, buildings and the indoor environment whereas outdoor positioning requires a regional or global coverage *(Mautz, 2008)*.

3.1.1 Outdoor Positioning

Outdoor Positioning solutions described in this section are the Global Positioning System *(GPS)* as one separate part of the Global Navigation Satellite Systems *(GNSS)*, a combination of GPS and metropolitan WiFi networks, WiFi fingerprinting methods and WiFi trilateration.
The GPS which is the most common technology supporting outdoor positioning *(Yeh et al., 2009; Li et al., 2008; Mautz, 2008)*. This satellite positioning system uses mainly the concept of 24 satellites in the earths' orbit sending radio frequency satellite signals. A GPS receiver on the ground needs at least 4 satellites for a

3D position and each satellite is the center of a circle. Further the Time of Arrival *(TOA)* is used to calculate the distance between the satellites. After this process a triangulation is carried out to find the intersection points of the circle. The Line of Sight between the GPS receiver and the GPS satellites is necessary to obtain an accuracy of approximately 5 meters to 40 meters, which is also depending on the number of satellites *(Yeh et al., 2009)* There is an average number of 10 GPS satellites available on the open sky *(Mautz, 2008)*.
GPS is one separate part from the GNSS which are used in a wide range of applications world wide. Current and future GNSS systems are GPS of the United States, GLONASS from Russia, Galileo from the European Union and Beidou from China. Nowadays, current GNSS systems are being improved and new systems are developed *(Mautz, 2008)*.
Satellites do not deliver a reliable position information in indoor environments and if buildings are within the line of sight. This problem can be seen in figure 3.1, where a person in the open field has a GPS connection, a person within a building not as there is no line of sight and between buildings where some satellites are blocked from buildings (Barnes et al., 2003).

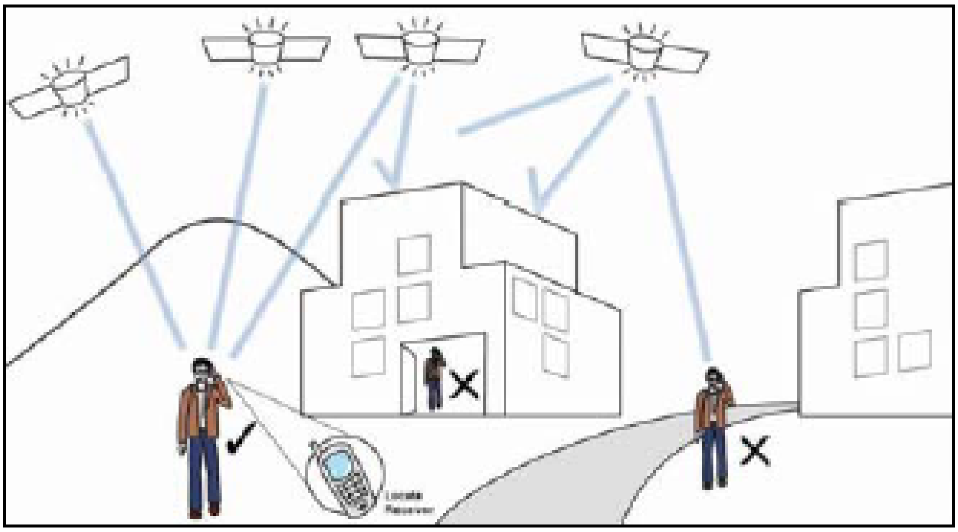


Figure 3.1: Positioning problems with GPS if the satellites are not in the line of sight (Barnes et al., 2003).

GPS has many different fields of application that start with electrical systems for timing or surveying methods up to logistic and traffic management. Other application fields are security, marketing, farming, banking, the car rental industry as well as weather forecasting *(Theiss et al., 2005)*.
Another opportunity for outdoor positioning is using a combination of the two technologies GPS and WiFi networks. This is possible through the development of the wireless Internet industry. Years ago the WiFi coverage was limited and it was a challenge to identify an accurate position. Nowadays, there is a rapid development WiFi coverage in metropolitan areas. For the study WiFi access points have been set up for positioning which are not disturbed by climate or line of sight as it is

for GPS, but there is a higher error rate. Finally, the implemented system is using portable devices to calculate the position using a pattern matching method of GPS and WiFi positioning. Weights are assessed to include the weather conditions for GPS if it is sunny or cloudy to increase the position accuracy *(Yeh et al., 2009)*.
Another outdoor positioning approach is the trilateration process, which requires at least three fixed base stations with referenced coordinates. A position can be determined using circles around each used reference position with a specific radius. All circles intersect at the position that has to be determined. The trilateration consists of two steps. The first step is regarding to a model establishing the distance of the radius. The second step is using geometric methods to compute the location *(Li et al., 2008)*.
Location based fingerprinting is based on measured radio signals as for example WiFi. As a human fingerprint, this location based fingerprint is associated with a location. This two phase approach starts with the training, where the fingerprints have to be established and stored into a database with characteristic features of the measurement. The second phase is the positioning itself where the measurements are compared with the reference points in the database and their characteristic features using searching and matching algorithms *(Li et al., 2008)*.

3.1.2 Indoor Positioning

Nowadays, indoor environment research is increasing as an average person spends about 90 percent inside a building and therefore indoor positioning is a crucial issue (Worboys, 2011; Giudice et al., 2010). This sub-section covers wireless LAN, Radio Frequency and Ultrasonic indoor positioning solutions. Indoor geography is also concerning severe limitations through these technologies as for example the position accuracy and further limitations through the size of rooms, buildings and the used context *(Xiang et al., 2004; Mautz, 2009)*. In general, many indoor positioning solutions are used for context-aware services on the location of a person or for example goods in a production line *(Xiang et al., 2004; Al Nuaimi and Kamel, 2011)*.

Wireless LAN based indoor positioning

The first described indoor positioning solution is the wireless LAN based indoor positioning technology. Many efforts are carried out to use the Wireless LAN network infrastructure within buildings, which is an economic solution (Xiang et al., 2004; Bill et al., 2004). Therefore, wireless LAN based positioning is a good approach for services that are limited in space such as specific areas of campus areas, class rooms, malls or other small indoor environments *(Bill et al., 2004)*. Trilateration or multilateration distance estimations are used in the indoor environment to calculate the position between the various wireless LAN access points *(Mautz, 2009)*.
To enhance an indoor based wireless LAN positioning the access points of the wireless LAN in a building are used. The access points are the basis for the determination of the requested position. The position determination process includes the

measurement of the signal strength, the angle of the arriving signal as well as the difference in time of the arriving signal. Problems occurring with wireless LAN based positioning are the limitation of the accuracy by the cell size and the effectiveness of the positioning by multiple reflections within rooms, furniture and equipment in rooms. An advantage of wireless LAN based positioning is simple coverage of a large area in respect to other indoor positioning solutions. Fixed positions can be determined with an accuracy of 2 meters and a moving position with an accuracy of 5 meters *(Xiang et al., 2004)*.

Indoor positioning based on the time of arrival or time difference

The second class of indoor positioning systems are using time of arrival or the time difference of arrival such as ultra sound systems or radio frequency systems. These systems are often combined and the time difference is used. Ultra sound pulses are sent together with radio frequency message that are synchronized with a unique identification from for example antenna lines on the top of the rooms. Through the measurements of sound and the time difference between radio frequency and ultra sound it is possible to determine the position by considering the fixed position of the antennas *(Mautz, 2009)*.
An example for an indoor ultra sound position system is the active bat system. Therefore, specific ultra sound tags are deployed on the ceiling and a listener is placed on a person. The active bat listener is calculating the position through the time of flight and a multilateration algorithm as several ultra sound sending tags are placed. It is also possible to determine the direction of the object *(Mautz, 2009; Koyuncu and Yang, 2010)*.

Radio Frequency Identification indoor positioning

Third, Radio Frequency Identification *(RFID)* is an indoor positioning solution which can work without the line of sight and no direct contact is necessary. This technology is working through active antennas that have to be placed in the indoor environment and tags receiving the signal. The position is determined by calculating the difference in the received signal strength of the tags and the antennas. An increasing accuracy can be achieved by placing more base stations or reference antennas in the indoor environment. For the active antenna a permanent power level is necessary. RFID based indoor positioning is a very accurate but cost intensive indoor positioning solution *(Koyuncu and Yang, 2010; Al Nuaimi and Kamel, 2011)*.
As it is shown there is a wide variety of indoor positioning solutions. This diversity is regarding to challenging demands of complex and high resolution building structures, the functionality of the building and the field of application which needs a specific accuracy *(Meijers et al., 2005; Ascraft, 2008)*. A comparison according to the accuracy and the coverage of indoor positioning solutions can be seen in figure 3.2. Mautz (2009) shows in this figure the diversity of indoor positioning solutions and their accuracy. Figure 3.2 is also enhanced with some outdoor positioning solutions that can also be adapted for indoor positioning.

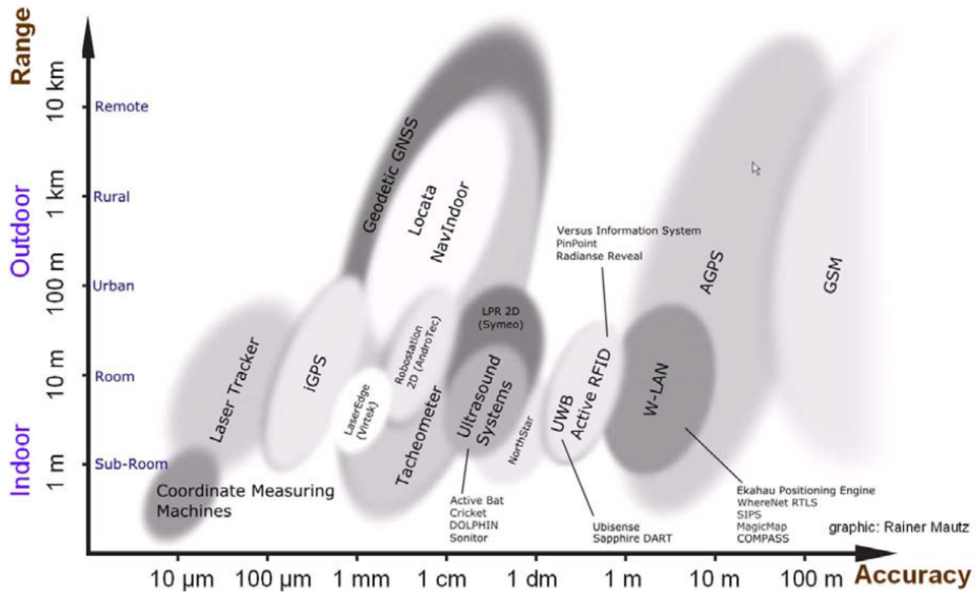


Figure 3.2: Accuracy and coverage of positioning solutions (Mautz, 2009).

3.1.3 Indoor Positioning and Production Lines

Production line environments are a very special indoor environment that have the task to support various production processes in an optimal way. Such a production indoor layout looks different than a building constructed for a residential or office use. A production line environment can have further constraints then the layout as for example a specific air quality or security aspects within a production line. Nyström et al. (2006) defines the production line as a challenging research field facing optimization regarding allocation and sequencing of production processes. A proper production line oriented for a higher customer orientation is often highly flexible with often changing arrangements of production units *(Thiesse et al., 2006)*. Main indoor positioning solutions for the comprehensive positioning tasks in a production line are radio frequency, infrared, ultrasound techniques and combinations of different technologies as described in sub-section 3.1.2 indoor positioning. The indoor positioning technologies implemented in a production line have special abilities for the specific environment including many reflected signals. Further is the number of objects which position is determined simultaneously. In a production line ceilings, walls and floors are often made of metal which makes the RFID communications more complicated and error prone. Another requirement for positioning within a production line environment is that there is not always a direct line of sight between the antenna and the receiver. The positioning alone is in a production line environment often too less. Another important aspect that can be accomplished is to ensure a direct communication with operators in the production line. To sum up, the requirements for indoor positioning in a production line are a high localization accuracy for a large number of production goods including a communication opportunity between the receivers and operators *(Thiesse et al., 2006)*.

Ultrasound systems are very accurate but they are not suitable for tracking many objects at the same time.*(Smith et al., 2004)*. Many production goods are often located in the same area, where the active bat system does not perform in a proper way *(Mautz, 2009; Koyuncu and Yang, 2010)*. According to Thiesse et al. (2006) the combination of RFID and Ultrasound is a good way to locate and to communicate in a production line environment.

3.2 Spatio-Temporal Data

Spatio-temporal data is gaining more and more importance as it is combining the position of an individual object changing over time *(Papadias et al., 2002)*. In outdoor geography spatio-temporal data can be for example traffic supervision, which is interesting to understand or manage the traffic flow, or pedestrian cell phones *(Papadias et al., 2002; Abraham and Sojan Lal, 2012)*. In indoor geography, especially companies spatio-temporal data is gathered storing several historical information about the movement behavior of production goods, to replay and reproduce states of the product at different historical steps *(Fan et al., 2011)*. Spatio-temporal data mining is an opportunity to figure out bottle necks, stops or candidate stops of production goods *(Alvares et al., 2009)*. Further spatio-temporal datasets are often large and spatio-temporal data mining helps to analyze and display the data *(Compieta et al., 2007)*. Therefore, a spatio-temporal database has to be developed including space and time in one model which is described in this section *(Fan et al., 2010; Lee et al., 2009)*.
Spatio-temporal data mining is a comprehensive and complex task as it is combing space and time which cannot be analyzed in a fully automatic way *(Andrienko et al., 2010)*. Many spatio-temporal applications are analyzing trajectories if they contain same patterns in a period of time *(Mamoulis et al., 2004)*. An example can be if a person is walking every day the same way to school or to work. Trajectories of moving objects have a high amount of gathered data that are in conflict with limited spatio-temporal querying capabilities *(Mamoulis et al., 2004)*. The visual analysis of spatio-temporal data is supported through Self-Organizing Maps *(SOM)* providing analytical perspectives. Clustering and reduction of dimensionality is also an important aspect of SOM *(Andrienko et al., 2010)*.
This section covers mainly the storage of spatio-temporal data in a database and therefore different approaches are described. Further are spatio-temporal data mining methods for trajectory pattern mining, SOM and analysis of the geographic and attribute space.

3.2.1 Storage of Spatio-temporal Data in a Database

The handling and storage of spatio-temporal data is not a straight forward task regarding to challenges of the different dimensions in space and time *(Pelekis et al., 2004; Fan et al., 2011)* The modeling of spatio-temporal data is regarding to tem-

poral semantics, spatial semantics and spatio-temporal semantics *(Pelekis et al., 2004)*.
Temporal semantics are for example the granularity of the anchored points on the time axis, temporal operations if something is within a specific period of time, the density if the time is discrete like timestamps or continuous, the order of the time and the lifespan *(Pelekis et al., 2004)*. Most important temporal aspects are the transaction and valid time, which are used to associate the spatial change with the time (Pelekis et al., 2004; Erwig et al., 1999). Transaction time is the time, when the event is recorded and stored in the database and valid time is the so called real-world time when an event happened (Pelekis et al., 2004; Erwig et al., 1999).
Spatial semantics are covering the structure of the space if it is raster or vector, the orientation or direction of the object in the real world, the measurement of an object to calculate for example the length and the topology if different objects are disjoint, overlapping, equal and so on *(Pelekis et al., 2004)*. The spatio-temporal semantics are depending on the data type that has to be adopted by each model, the primitive notions specifying the abstracted data model, the evolution in time and space describing the movement or change of objects in space, space-time topology if a model has the ability to estimate metrics, the object identities and the dimensions *(Pelekis et al., 2004)*. Another point of the spatio-temporal semantic is highlighting the type of change of the object which can be seen in figure 3.3 pointing out the eight different possibilities of a spatio-temporal object (Pelekis et al., 2004).

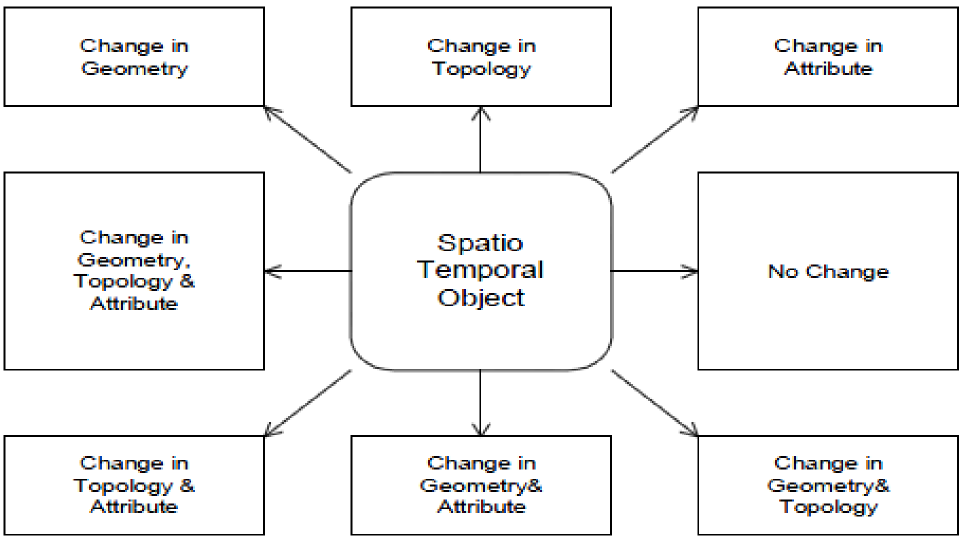


Figure 3.3: A spatio-temporal object and the types of change in the spatio-temporal semantic (Pelekis et al., 2004).

Spatio-temporal databases have different models as background. These models are for instance a snapshot model, a model based on timestamps and an event-oriented model that are used for different tasks and provide different query capabilities*(Pelekis et al., 2004; Erwig et al., 1999; Schneider, 2009)*.

Snapshot Model

The spatio-temporal snapshot data model involves timestamps separately for each layer and every object is a collection of homogeneous objects. This snapshot approach establishes a temporal attribute of the location and is the simplest way to visualize and store spatio-temporal information. To identify changes of an object in time, snapshots have to be compared. A so called snapshot includes at each timestamp every information, so unchanged objects and states are duplicated and only the valid time is supported. To sum up the snapshot model enables simple spatio-temporal queries and shows the spatial distribution at different times without a clearly defined temporal relation *(Pelekis et al., 2004)*. Problems can occur through the appearance of redundant entries if there are unchanged states of objects.

Data Model based on Time-Stamping

This is a simple approach tracking the state of an object with a tuple of timestamps, where the timestamps are the start and the end of a change of an objects' state. The end time can be defined as a timestamp or as a state or period of time. An advantage of this approach is to identify the state of an object at a specific time. The disadvantage is the missing capability to directly obtain information of what happened. So the effect of the change can be detected but not how it changed *(Pelekis et al., 2004)*. An example can be the tracking of a good in a production line. The tracked positions can be visualized, but it is not known what happened in between the tracking. In this case, the trajectory is an approximation of the real path between each of the tracked positions. The granularity and frequency of tracked positions have an effect on the trajectories, as if there is a higher frequency the trajectory is a more exact assumption.

Event-Oriented Models

An event-oriented model identifies each change or event of an object. Therefore, every event has to be stored individually. Pelekis et al. (2004) describe the event-oriented data model as logging of every change in a traditional GIS. The logged information includes everything for a spatio-temporal system. The logging is an association of every event with a timestamp and an indication where the change occurred. The current status can be seen as well as every state in the past. Advantages of this spatio-temporal data model are no redundancy of not changing objects and complex queries are supported such as what happened in a specific area in a period of time *(Pelekis et al., 2004)*.
To sum up, the storage of spatio-temporal data is not a straight forward task involving temporal semantics, spatial semantics and spatio-temporal semantics. Different spatio-temporal data models can be implemented. Therefore, the complexity of querying has to be considered as well as the availability of data.

3.2.2 Trajectory Pattern Mining

In recent years the moving-object trajectory data is increasing, which provides a fundamental issue for new techniques to analyze trajectories, similar groups of trajectories and to understand the behavior of the moving-objects *(Jeung et al., 2011; Alvares et al., 2007)*. Nowadays, mobile devices, smart-phones, navigation devices provide the opportunity to capture position information at a certain interval and can be combined to a trajectory *(Alvares et al., 2007)*. Trajectory patterns are relative motion patterns enabling relationships of motion attributes, disc-based and density-based patterns covered by *Jeung et al. (2011)*. Basic applications for trajectory pattern mining can be transport optimization, analysis of traffic and movement analysis of humans and animals *(Ben Aoun et al., 2014; Jeung et al., 2011)*.

The process of pattern discovery is described by Jeung et al. (2011) in figure 3.4. In the figure it can be seen that an object is prepared with a positioning device with a unique identifier. As the real trajectory is unknown, the tracked positions provide the opportunity to create an approximation of the objects' trajectory. Current techniques for trajectory pattern mining use polylines or a sequence of connected lines as a trajectory. The data collection is the recording of positions to a server in real time or if a device is offline a batch can be send later on. The collected trajectories have to be pre-processed and sampled. The server is somehow cleaning the data and prepares it for the trajectory pattern mining. The data format can either be stored as a sequence of positions or by a sequence of linear functions. The refined trajectories are later on used for trajectory pattern mining *(Jeung et al., 2011)*.

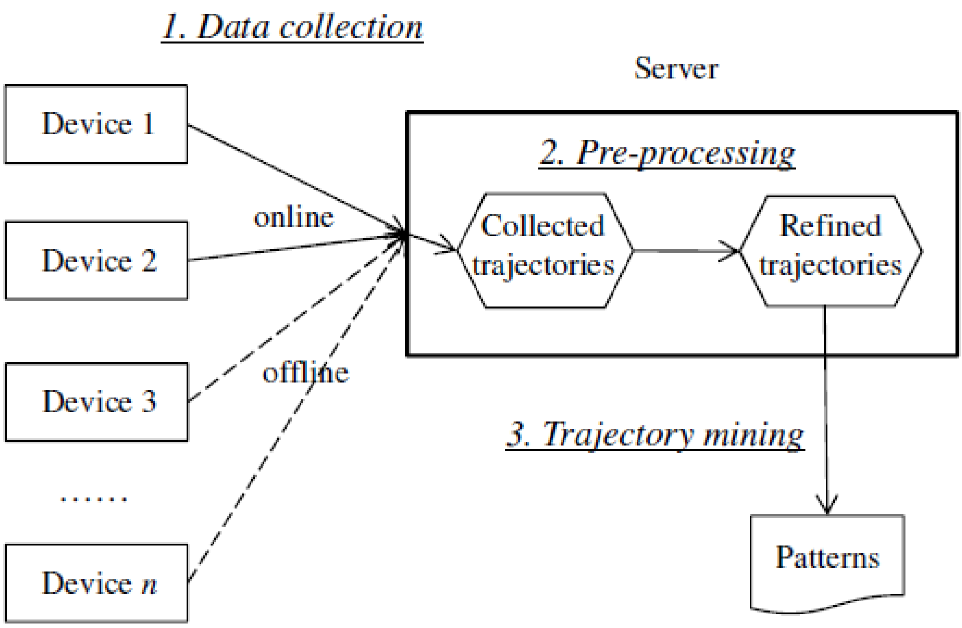


Figure 3.4: Trajectory discovery process. (Jeung et al., 2011).

The key features of relative motion patterns identify similarities of moving-object trajectories based on motion attributes such as speed or the azimuth of the motion. Relative motion pattern types are basic motions like constance, concurrence and

trendsetter. A constance motion is an equal motion attribute at an ongoing time and a concurrence is if multiple objects have the same motion attribute. In the case of a motion pattern which is shared by other objects in the future is a trendsetter. The motion attributes can be compared in a matrix of time and the trajectories. Another type of relative motion patterns are spatial motion patterns such as leadership, which is a trendsetter in a certain spatial constraint. Flock is a set of moving objects sharing the same motion attribute, which is a concurrence combined with a spatial constraint. The last spatial motion is a track, an individual object keeping the same motion attribute over a spatial unit. Further motion attributes are convergence describing a set of moving-objects sharing the same motion azimuth at a given spatial range. An encounter is a future meeting of a set of moving objects within a defined spatial range. A breakup is the opposite of an encounter and a divergence is the opposite of a convergence *(Jeung et al., 2011)*.
Disc-Based Trajectory Patterns are an extension to the relative motion patterns. Euclidean distances define the patterns instead of motion attributes. Spatial constraints are used as well as the integration of time. A prospective pattern is analyzing future trajectories under the assumption of the same speed and direction of the moving object. Therefore, an encounter is describing a group of a number of objects that will arrive in a disc with a predefined radius, as well as the number of objects has to be defined. A convergence is also describing a predefined number of objects passing through a defined disc. Flock-Driven patterns like a flock define a group of a defined minimum number of objects moving together at an ongoing time while staying in a disc with a defined radius. Meet is another flock-driven pattern including a group of a minimum number of objects staying together in a fixed disc for an ongoing time. Figure 3.5 shows a flock on the left where three moving objects are in the same area for a period of time and on the right a meet of 2 objects in a fixed disc. The flock and the meet need three parameters that have to be defined, the minimum number of objects, the disc radius and the time. Problems can occur by defining the size of disc, as a too big radius involves objects that are not in the same group and a too small radius is loosing objects of a group *(Jeung et al., 2011)*.

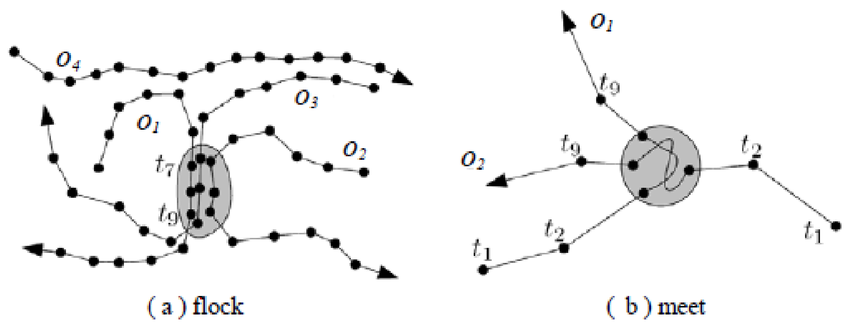
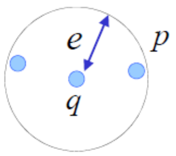


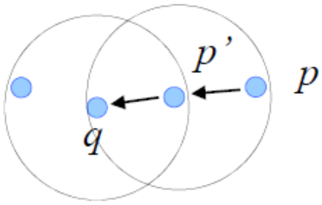
Figure 3.5: Flock-Driven Patterns of trajectories. (Jeung et al., 2011).

To capture and analyze trajectory patterns with an arbitrary shape and extend density-based trajectory patterns are established with distance related conditions.

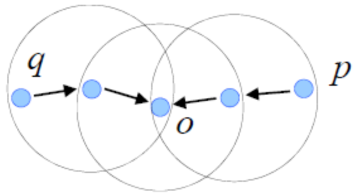
Basically density-based trajectory patterns take three different states into account, if a point is direct density reachable, density reachable or density connected which can be seen in figure 3.6. Density notions can be defined using two parameters *e* and *m*. A direct density reachable notion describes if a point p is directly density reachable from point q in respect to a distance e and a number of points m can be seen in sub-figure 3.6a. Density reachable is shown in sub-figure 3.6b, where a point p is density reachable from another point q under the distance constraint e and the number of points m if a chain of points exists between point p and q. The third density notion is density-connected and can be seen in sub-figure 3.6c. A point p of a set of points is density connected to a point q of the same set of points, if another point is existing which is density reachable from point p and point q under the distance constraint e and m *(Jeung et al., 2011)*.



(a) Direct Density Reachable.



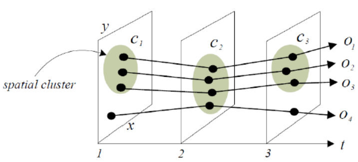
(b) Density Reachable.



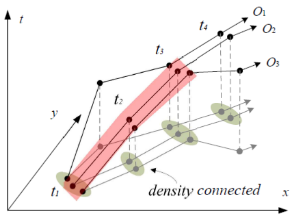
(c) Density Connected.

Figure 3.6: Density notions for trajectory pattern mining where the parameter m = 3 *(Jeung et al., 2011)*.

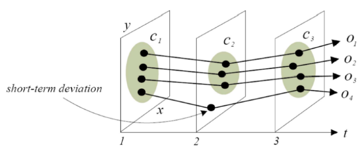
A moving cluster is a set of tracked objects that move close together of a consecutive period time. Sub-figure 3.7a shows a spatial cluster in an ongoing period of time. A convoy is defined to be density connected objects through an ongoing time interval, which can be seen in sub-figure 3.7b in red. Such a convoy exists of three parameters e distance, m number of points or objects and k for a minimum number of ongoing timestamps. The last described moving cluster is a swarm which can be described as a convoy without an ongoing period of time. The three parameters e, m and k have to be defined, whereas the time points do not have to be consecutive. In sub-figure 3.7c an example for a swarm is shown, where at the second time point one moving object is outside the cluster and goes back into the cluster at the third time point *(Jeung et al., 2011)*.



(a) A Moving Cluster.



(b) A convoy.



(c) A Swarm.

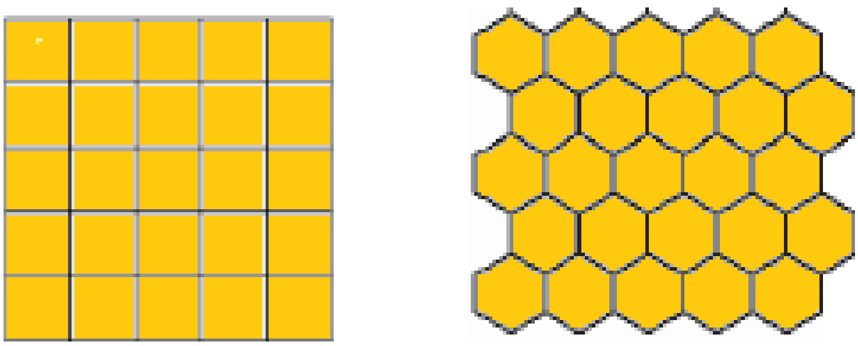
Figure 3.7: Moving Object Clustering opportunities like moving clusters, convoy and swarm *(Jeung et al., 2011)*.

3.2.3 Self-organizing Maps

A Self-Organizing Map *(SOM)* algorithm is one type of a neural network algorithm, which is an automatic data analysis method *(Kohonen, 2012)*. There are no logical or semantic relations between the neurons, only the distance to each other neuron or the neighborhood is known. It is a new and effective opportunity to visualize high-dimensional distribution over a low-dimensional grid *(Kohonen, 1998; Kohonen et al., 1996)*. Further, a SOM is compressing information but still takes topological and metric relationships of the base data on the display and leads to an abstraction of the data *(Kohonen, 1998)*. According to Kohonen (1998), this different type of visualization and abstraction can be applied to solve complex tasks like process analysis, perception of machines, control and communication. Summarizing SOMs are used for the visualization of data and visual data mining.
After a general description of SOMs, one possible training approach is described. Further this section covers visualization opportunities of SOMs and applications.

What's a Self-Organizing Map?

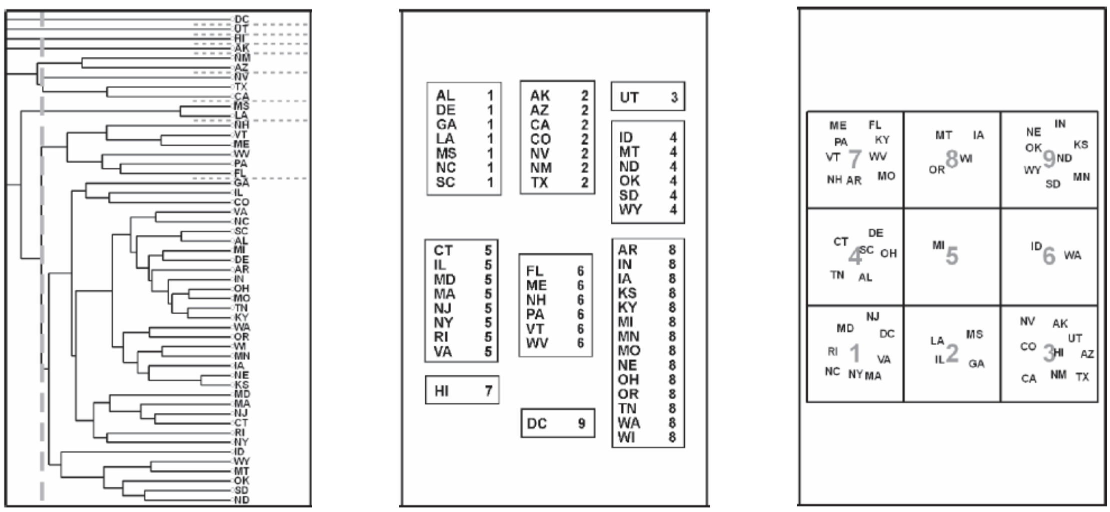
Self-Organizing Map, Self-Organizing Feature Map or Kohonen Map are all terms for the same type of map developed by Teuvo Kohonen and are regarding to one type of artificial neural network *(Kohonen, 1998)*. Unsupervised learning methods are used to train the map, which is regarding to the term "Self-organizing". SOMs are commonly represented as a two-dimensional grid existing of so called neurons, which are existing in two different shapes, rectangular or hexagonal. The topology of a SOM affects the neighbors of neurons. In a rectangular topology each neuron has a maximum of four direct neighbors, whereas the hexagonal shape results in a maximum of six neighbors for each neuron. Further the hexagonal representation has a smoother visualization and is commonly used. Figure 3.8 shows both visualizations in figure 3.8a the rectangular topology and in figure 3.8b the more frequently used hexagonal topology.



(a) Rectangular topology of a SOM. **(b)** Hexagonal topology of a SOM.

Figure 3.8: A two dimensional SOM topology displayed as a rectangular and a hexagonal arrangement.

Often a SOM is described as a clustering method or some method of dimensionality reduction of an n-dimensional space. On one hand, researches dealing with clustering methods such as *k*-means clustering identify parts of this clustering or classification method in a SOM. Additionally, the SOM provides a particular representation including also topological capabilities of the clusters. On the other hand, researchers covering topics with data in *n*-dimensions will recognize aspects of a principal component analysis or multi dimensional scaling in a SOM. Agarwal and Skupin (2008) show in one example the difference between a hierarchical clustering method, *k*-means clustering and a SOM. Therefore, the similarities or dissimilarities of attributes are used to investigate clusters or groups of data. The used data for the example is regarding to the 50 states of the US including 32 attributes for each state. Figure 3.9 shows three different clustering methods the common traditional hierarchical clustering methods of the similarities, *k*-means clustering and the clustering using a SOM. In figure 3.9a a dendrogram is visualizing the traditional hierarchical clustering where the tree structure is implemented by splitting all states into smaller groups until each state is alone or vice versa. Compared to this approach the *k*-means clustering in figure 3.9b investigates the data by creating *k* - partitions of the data. In this case the *k* = 9, resulting in 9 clusters. Additionally, figure 3.9c shows the result of 3 * 3 SOM in a two-dimensional rectangular shape. The advantage of the SOM approach is, that the solution is in a topological order. It can clearly be said that there is a high difference of the attributes behind the cluster 1 and cluster 9 as they are no neighbors. In contrary cluster 1 and cluster 2 are neighbors which shows that the clusters have overall also some similarities between each other *(Agarwal and Skupin, 2008)*.



(a) Hierarchical Clustering. **(b)** *k*-means clustering. **(c)** SOM.

Figure 3.9: Difference of clustering methods on an example of the 50 US states and demographic data (Agarwal and Skupin, 2008).

Agarwal and Skupin (2008) applied the same example on the demographic data of the 50 US states with a principal component analysis, multi dimensional scaling and a SOM to show the difference. This is not described in this thesis, as it is giving a general overview of SOM.

Stepwise training Approach

A typical SOM algorithm consists of six base steps the data preparation, neuron initialization, stepwise SOM training, neighborhood functions, batch computation of the SOM and the end results of the SOM training process. After the training process of the neurons and the association process of the best matching neuron it can be visualized as a SOM *(Kohonen, 1998, 2012)*.

First is the data preparation phase including the preparation of quantitative data to enable a comparison of the values, as well as each value has to stand alone and the data that has to be investigated must be complete *(Kohonen, 1998)*. Second, the task is to initialize the neurons the decision about the number of neurons has to be answered, which has a great effect on the computing performance to train the neurons *(Vesanto et al., 2000)*. The arrangement of the neurons has also to be selected as there is the opportunity of for example a hexagonal shape or a rectangular shape. The rectangular shape has advantages with neighborhood relations, because of the symmetrical orthogonal arrangement *(Vesanto et al., 2000)*. According to Kohonen (2012),the neurons are often initialized with a random value in the range of the input data to decrease the training steps. The last step for the initialization process is the selection of the overall shape which is most of the time rectangular or quadratic.

After the initializing of the neurons the third phase is the stepwise training of the SOM starts. Therefore, the input vectors are compared with all the neurons to find the matching neuron with the smallest Euclidian distance to the specific input data, the so called best matching unit *(BMU)*. Figure 3.10 illustrates the stepwise SOM training process. An input feature is given as *X* and compared to each initialized neuron *M* and the stepwise comparison is visualized, till the smallest Euclidian distance is found. After this searching process, the neighbor neurons are modified and get more similar to the input vector. By an increasing distance to the matched neuron for the input vector the neurons are less modified which is a decreasing in the learning rate. The process ends if there is no learning rate anymore and the next input vector has to be compared. If this step is executed for each input vector the neurons of the SOM are spatially ordered and represent the input data *(Kohonen, 2012)*.

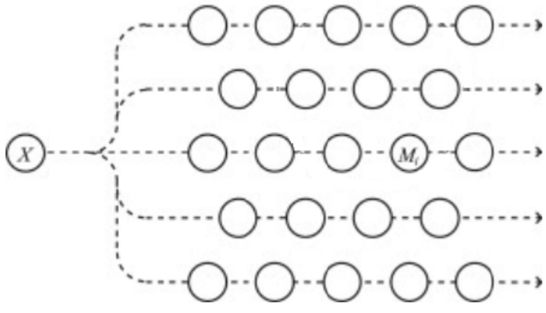
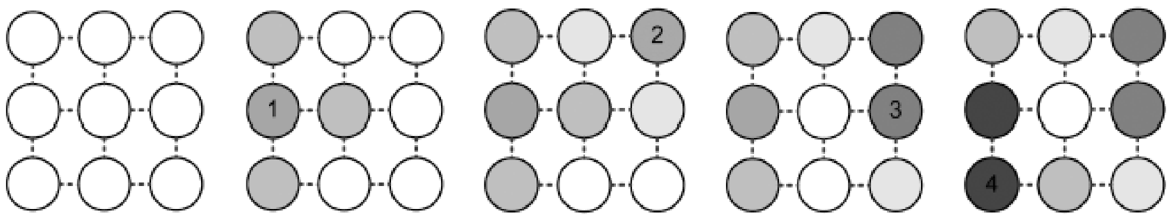


Figure 3.10: A specific input vector *X* compared to all neurons *Mj* to find the best matching input vector.

Fourth is the neighborhood function to describe the decreasing learning rate of the neighboring neurons with an increasing distance to the best matching neuron. Therefore, a so called neighborhood function is used as for example the Gaussian neighborhood function $h_{c(x)i}$:

$$h_{c(x),i} = \alpha(t)exp\left(-\frac{\|r_i - r_c\|^2}{2\sigma^2(t)}\right)$$

In this Gaussian neighborhood function, $0 < \alpha < 1$ is the decreasing learning rate factor according to the number of *t* training steps. The vectors r_i and r_c cover the positions of the selected neuron and the best matching neuron. Also decreasing with the number of training steps is $\alpha(t)$ representing the width of the neighborhood function *(Vesanto et al., 2000)*.
An abstract example of the training approach is given by Agarwal and Skupin (2008) with a 3 * 3 SOM in figure 3.11. At the initial point of the training of the neurons every neuron has a random or is empty as it can be seen in figure 3.11a. The training starts with the first vector and the BMU is identified and through the neighborhood function the other neurons are adjusted. This step can be seen in figure 3.11b, number 1 shows the BMU and in a lighter gray the neighboring neurons are adjusted. The same step is carried out for the second input vector in figure 3.11c in another gray scale with the number 2 as the BMU and the adjusted neighboring neurons. Figure 3.11d shows how the third input vector is matched to the BMU which is next to the second input vector from before and the effect of adjusting the neighboring neurons can be seen again as the neurons have a different gray scale. Finally figure 3.11e shows how the fourth and last input vector is matched and the neurons are adjusted accordingly. The result of the described training processes is a 3 * 3 SOM with four trained input vectors.



(a) Before the training. **(b)** First vector is trained. **(c)** Second vector is trained. **(d)** Third vector is trained. **(e)** Fourth vector is trained.

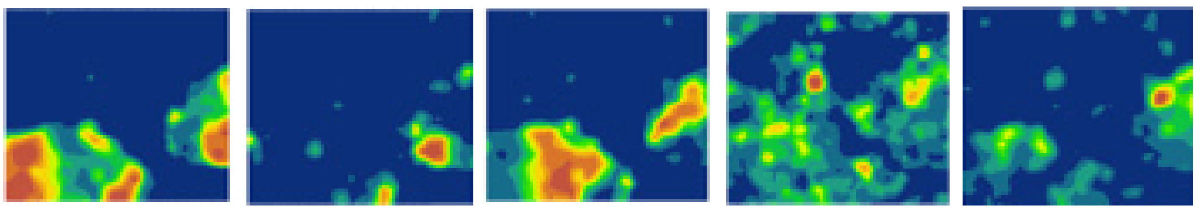
Figure 3.11: Schematic training of a 3 * 3 SOM. Different groups are trained and visualized in a gray color schema (Agarwal and Skupin, 2008).

Fifth is another training opportunity for a SOM, the batch computation. This approach has less computational effort compared to the stepwise training described in step 3 before. The initialization process of the neurons is equally as well as the identifying of the BMU of the neurons. Basically, each input vector is matched to the neurons and the best matching unit is found without modifying the neighboring neurons. The vector values are stored in a sub-list which is linked to the best matching unit. Finally, for each neuron the average of all linked values is calculated

and weighted with a neighborhood function *(Kohonen, 1998)*.
Sixth is the end result of the SOM training and the visualization. The training process of a SOM ends after a predefined number of steps or circles and in some cases a certain threshold of change per training step can be defined. The number of training steps depends on the quality of input data, the intended usage *(Kohonen, 2012)*. There is also the opportunity to project the result of the input data over a geographical unit to identify similarities or spatial differences.

Visualization of SOM

Basically the visualization techniques used for SOMs are covered by Skupin and Esperbé (2011) describing the SOM itself, the mapping of n-dimensional vectors onto a SOM and the linking of a SOM with other display spaces. As a part of this thesis only some examples are taken into account giving an overview of the capabilities of a SOM. A combination of a SOM and a geographic map,the SOM grid, trajectories and component charts projected onto a SOM as well as distance-preserving visualization and a parallel coordinate plot projected from a SOM are described.
SOM Grid is a common visualization method splitting up the SOM into component Planes. In this case, a component plane is the mapping of the values of one attribute over the SOM. The component planes are increasing with the number of attributes a user displays. This visualization technique provides the capability of investigating correlations between the various components of a SOM. Therefore, a visual inspection of the component planes is possible *(Rainer, 2013)*. An example of such a component plane is given by Skupin and Esperbé (2011) in a high-resolution 500 * 500 SOM with a high number of attributes such as climate, soil, population and many more. Figure 3.12 shows an extract of five selected component planes. By visual inspection it can be seen, that in the left corner of the agricultural component plane the red values indicate high values of agriculture. It can also be seen that in this agricultural cluster another type of land use is present. Therefore, the forest land component plane shows that the forest cluster in the left bottom fits in this space of the agricultural component plane.



(a) Agricultural. **(b)** Rangeland. **(c)** Forest Land. **(d)** Water. **(e)** Wetland.

Figure 3.12: An extract of five component planes of a 500*500 high-resoltion SOM (Skupin and Esperbé, 2011).

Remark: Further techniques to visualize the SOM itself are listed and described by Rainer (2013). These techniques are for example Vector Fields, Cluster Connections,distortion patterns, U-Matrix or other Distance Matrices.

Mapping data onto a SOM is another opportunity to explore the created neural network after the whole training processes. A trajectory can be created by the usage of the changing location of the calculated BMU after several training steps. Therefore, Agarwal and Skupin (2008) applied such a trajectory in a research. Basically the change of three states of the US are mapped on a SOM after every 10.000 iteration of the training process and the time-shift is creating the trajectories. In other words, the trajectory of the BMU of each of the three states New Jersey, Illinois and Minnesota is mapped which can be seen in figure 3.13. It shows that after the random initialization the two US states Illinois and Minnesota look quite similar in a high-dimensional space, but after each training run they move further apart. Minnesota is directly moving down into the right corner of the SOM going further up in the right corner until it comes to an end. In contrary, the trajectory of Illinois goes directly up in the top left corner of the SOM and returns more in the middle of the SOM *(Agarwal and Skupin, 2008)*.

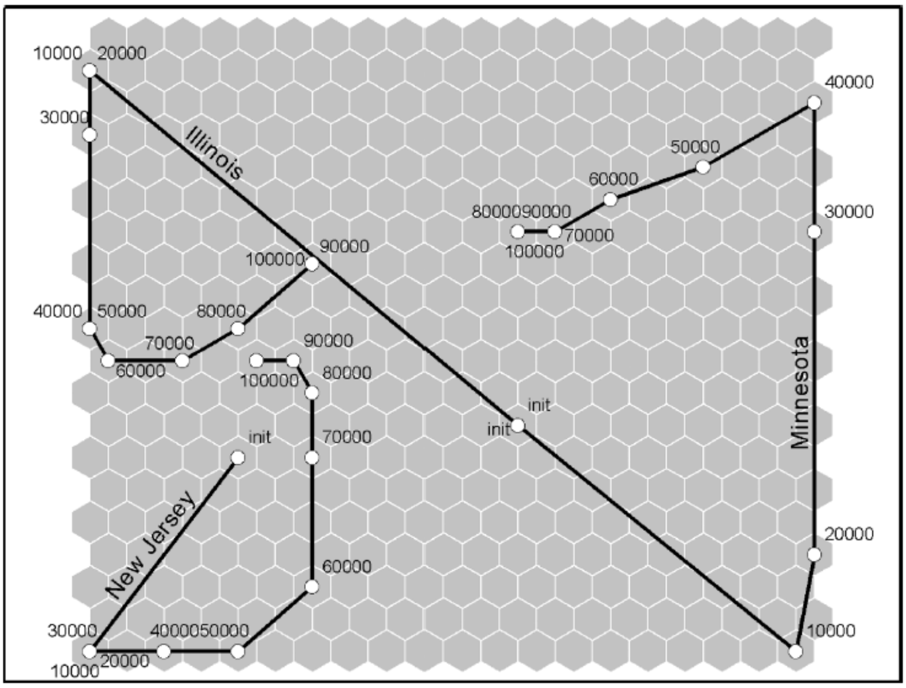


Figure 3.13: Change of the BMUs position of three US states after every 10.000 training iterations (Agarwal and Skupin, 2008)..

Remark: Further techniques to map data onto a SOM are listed and described by Rainer (2013). These described techniques cover data histograms, component charts projected onto the SOM, neighborhood graphs such as the Minimum Spanning Tree and a class map.

SOM linked with Geographic Maps is an important representation as there is already a visual form existing for the user and the interpretation of the map. The connection between a SOM and a geographic map is basically done by the usage of the BMU of the *n*-dimensional SOM data, which is mapped to the geographic space. Through the different projections of the data affected areas can easily be

located on the map. Such a geographic map resulting of SOM data is the basis for a further geospatial data analysis as the results of the n-dimensional attribute space are linked to a geographic representation *(Agarwal and Skupin, 2008)*.
To demonstrate this visualization technique of a SOM, Skupin and Esperbé (2011) created a high-dimensional SOM and executed the k-means clustering on the SOM to identify clusters on the SOM. These cluster are mapped onto the geographic space of the corresponding data. In this case attributes of US census block groups are analyzed and the US census block groups are geographically mapped, which can be seen in figure 3.14. The clusters are numbered accordingly. Some clusters from the SOM seem not to appear on the geographic space at this scale, as they are only a few US census block groups and sometimes also located in cities.

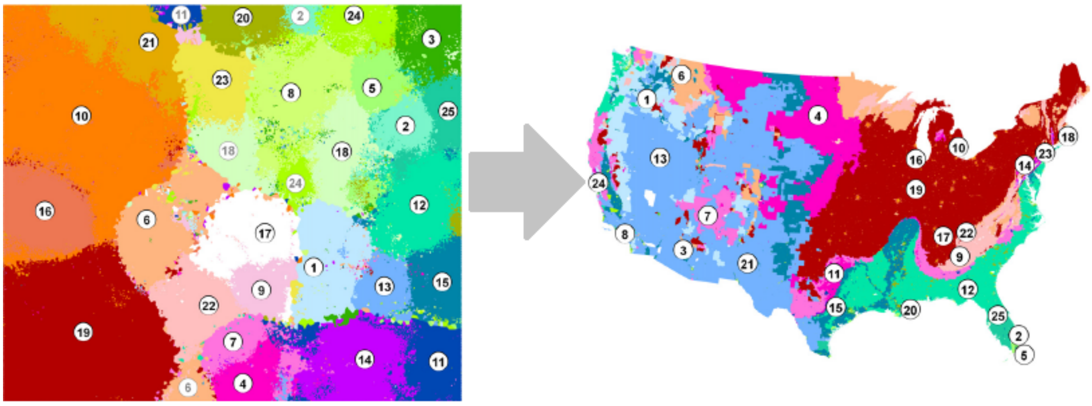


Figure 3.14: Mapping of SOM onto the corresponding geographic space (Skupin and Esperbé, 2011).

SOMatic Software Application

A stand-alone software was developed in 2013 separated into two parts with the aim to find patterns in high-dimensional datasets. On one hand the SOMatic Trainer developed by Spöcklberger (2013) and on the other hand a SOMatic Viewer developed by Rainer (2013) which was part of their Master Thesis. In general, SOMatic is developed using the programming language *JAVA* and additionally *PROCESSING*, which is a programming language and development environment.
The **SOMatic Trainer** mainly focuses on three steps the data pre-processing, the initialization process of the SOM and finally the SOM training. A graphical user interface is accessing the SOMatic Trainer Library carrying out these three steps. Therefore, the input files have to be either a Comma Separated Value file format **.csv* or a **.dat* file. Settings for the pre-processing steps can be selected in the SOMatic trainer user interface as well as the initialization of the SOM and the training steps of the SOM. After the training process the SOM files **.cod* (Codebook) and the **.sprj* (Project File) can be stored locally. Additionally, the library provides the possibility to visualize an animation of the training process *(Spöcklberger, 2013)*.

The **SOMatic Viewer** is mainly implemented to visualize SOM in an interactive way. The knowledge gaining process is described by figure 3.15.The training process of a SOM is the first task which can be seen on the left site and is in this case done by the SOMatic trainer. It is also possible to pre-process and train the SOM with other software tools, as long as the output files are a **.cod* and **.dat* file. These files are loaded into the SOMatic Viewer and the features are extracted for a further visualization. The visualization itself needs the data from the SOM as well as external input data for the geometry as **.SHP* file. In the end is the visual data exploration and the knowledge gaining process. Therefore, the Shapefile and the before pre-processed data from the SOMatic trainer need the same unique identifier to map the features properly and enables different visual outputs *(Rainer, 2013)*.

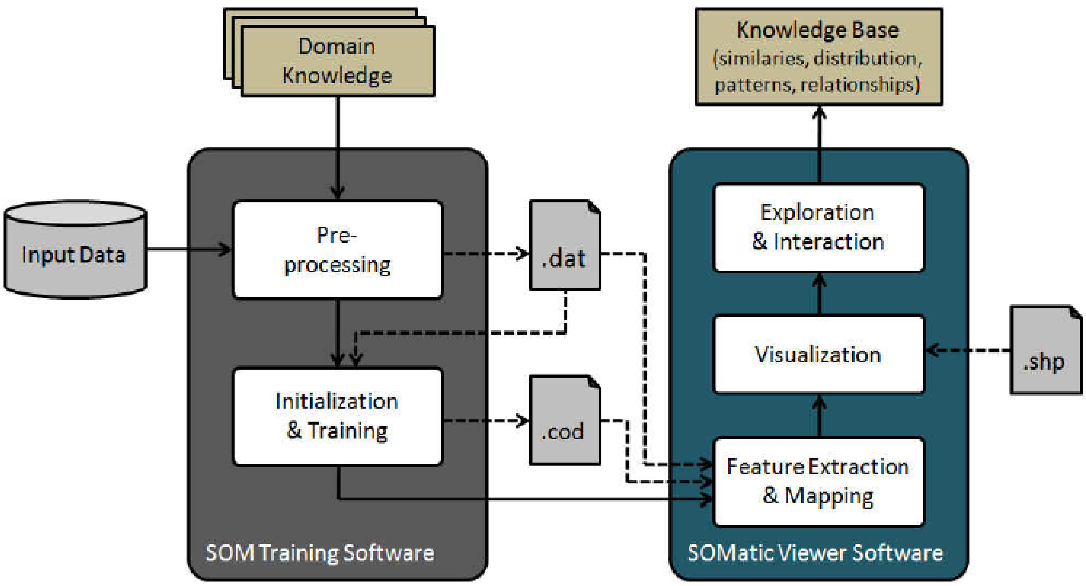


Figure 3.15: Knowledge gaining process workflow including the SOMatic trainer and the SOMatic viewer (Rainer, 2013).

3.2.4 TRI-Space Approach: Location, Time and Attributes

The geographic space, temporal space and attribute space is the so called TRI-space and they are investigated in spatio-temporal datasets *(Skupin, 2010)*. An example for such a tri-space can be information of a specific geographic information at a particular timestamp, like a temperature measurement at a geographic location at 6 o'clock in the morning. Figure 3.16 shows the tri-space elements occurring in a spatio-temporal dataset and each space the geographical, the attributes and the time are represented by one or more instances such as one attribute of the dataset *(Skupin, 2010)*. According to Skupin and Esperbé (2011) features in geographic space and high-resolution attribute space can be created through SOMs and geographic features can also be projected into the attribute space.

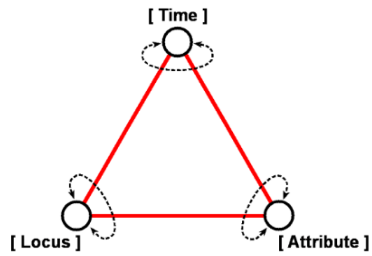


Figure 3.16: Tri-space of location, attributes and the time (Skupin, 2010).

In general, geographic features are visualized at a very high level of detail, while the attributes and relationships of the features are included at a sparse level limiting the understanding of higher dimensional data *(Skupin and Esperbé, 2011)*. The tri-space includes six constructed spaces from either a single element as location, attribute and time or combinations of two elements like location and time, time and attribute or location and attribute *(Skupin, 2010)*. Therefore, the triangle in figure 3.16 represents single elements through the corners and the combinations of spaces are represented as edges. Challenges are regarding to the transformation between different TRI-space perspectives, in particular if there are missing values. In addition, to analyze the high-dimensional data there should be a normalization or a weighting before the processing steps *(Skupin, 2010)*.
According to Skupin (2010), one of the six spaces is describing a combination of time and attributes linked to an individual location. This is the so called **L-AT** perspective. As most of the base data for a conceptualization are in a tabular form, this **L-AT** perspective is described in a tabular form and leads to a spatialization in a 2D map. In this tabular form the location regards to the rows of the table and the attributes and the time combination represents up the columns of the tabular input data. This **L-AT** perspective is the basis for questions about multi-temporal, multi-attribute relationships of locations. Figure 3.17 represents on the left the L-AT perspective from the TRI-space triangle from figure 3.16. The middle highlights the tabular form and finally in the right is the spatialization in 2D *(Skupin, 2010)*.

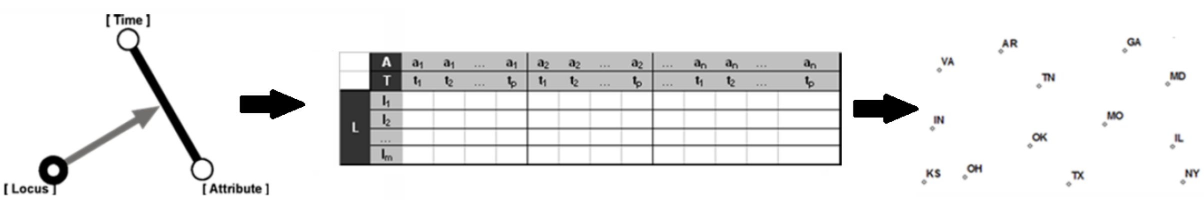


Figure 3.17: L-AT perspective of the TRI-space starts with a basic conceptualization followed by the tabular structure of the data and a spatialization in the end (Skupin, 2010).

By taking into account the **LT-A** perspective of the TRI-space, a location and time combination is used as objects which is referring to domain specific attributes. This perspective of the TRI-space provides more detailed questions of temporal aspects of locations. Trajectories of locations over time enable the exploration of location trajectories over time. Figure 3.18 shows the main concept of the **LT-A** perspective on top followed by the tabular form of the input data in the middle which leads to a trajectory visualization in the bottom. The trajectory visualization shows, that the two green location-time objects have similar attributes and contrary the two red trajectories are not similar over time *(Skupin, 2010)*.

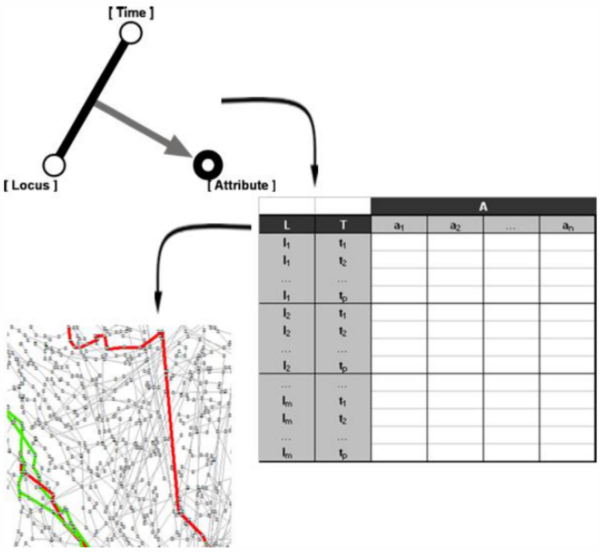


Figure 3.18: LT-A perspective showing the steps from the conceptualization, tabular input data and a spatialization. In this case, the green track describes similar objects and the red tracks not similar objects (Skupin, 2010).

3.3 Routing in the Indoor Environment

Different approaches enabling the routing indoor environment use the topology as a basis *(Raubal and Worboys, 1999)*. In addition, the indoor space is established or reduced to a graph *(Goetz and Zipf, 2011; Goetz, 2012)*. An example for such an approach with a graph is established by Jensen et al. (2009) as they implemented a graph based model to track several entities in an indoor environment using sensors distributed in the indoor environment. Building Information Systems (BIMs) are used to model the 3D geometry of buildings. In general, BIMs do not support navigation and routing abilities *(Howell and Batcheler, 2005)*. Whether, there are approaches showing BIM oriented models supporting specific indoor navigation requirements.

One BIM approach is developed by Isikdag et al. (2013) and determines the problem as there is a lack of sufficient approaches of existing indoor navigation based on 2D geometries and pre-defined routing abilities for emergency response, delivery, utility maintenance and facility management. These are all key issues for companies with a production. To overcome this insufficiency, a BIM provides advanced semantic and geometric information, which is a very complex task to bring the BIM into a structure facilitating the indoor navigation. The approach of Isikdag

et al. (2013) provides highly detailed semantic information for routing purposes as well as the representation of non-geo-referenced structures and the various complex structures of a BIM. Their BIM Oriented Modeling approach is called BO-IDM, a BIM based model enabling the navigation in the indoor environment *(Isikdag et al., 2013)*. According to Volk et al. (2014) the automated creation and updates of BIM is a challenging task. Further, there is a rapid development of BIMs especially for new buildings, whereas there is a lack of BIMs for most of already existing buildings (Volk et al., 2014).
Hybrid models for indoor navigation including geometrical and topological features are mentioned by Worboys (2011). These hybrid models are also well studied in recent literature *(Stoffel et al., 2008; Becker et al., 2009)*. Further approaches focus on different levels of granularity, the user has the ability to rely on details for important points on a journey ,such as the beginning, end or any decision point where there is a user interaction, resulting in the generation of a route and the visualization *(Hagedorn et al., 2009; Stoffel et al., 2008)*.

3.4 Ontology for the Indoor Environment

To model the indoor environment of a production line accordingly, ontologies are used describing the indoor environment in a simple and formal way. In general, an ontology has the aim to identify the types and categories of objects and their relations in a generic way *(Smith, 2001)*. Raubal (2001) states that an ontology can also be domain specific, in this case the ontology describes the specific domain in a generic way. Therefore, a formal description including the whole content of the specific domain and the behavior has to be described as it is a part of the physical world *(Raubal, 2001)*. The elements of an ontology are listed by Davis (1990), which are the entities, the relations of entities and rules that have to be applied on the entities and the entities' behavior.
The indoor environment of a production line has many specific constraints and limitations according to air quality, security issues, production optimization, routing, etc. Therefore, the theory of an affordance based model of the indoor environment is considered to model the routing of production assets between the production line processes, determining the best possible path according to the stated constraints. The term affordance is mainly developed by Gibson (1977). Recent research in outdoor and indoor environments is focusing on affordances and ontoligies *(Turner and Penn, 2002; Kapadia et al., 2009; Kim et al., 2012)*. Anagnostopoulos et al. (2005) and Tsetsos et al. (2006) developed an ontology focusing on the navigation in an indoor environment. Further, Yang and Worboys (2011) developed an ontology for the navigation in the indoor and outdoor environment. In their research, they distinguish between different "microworlds" by considering the upper level ontology, a domain specific ontology and a task oriented ontology.

Chapter 4

Methodology

This chapter covers the methodology for the proposed approach. It includes a conceptional workflow model of the overall approach to highlight the main steps of the methodology and the implementation. This is followed by a delineation of the production line under review and the indoor positioning solution used in the examined production line. The modeling phase is pointing out the definition of the use cases, the spatio-temporal data concept as well as an abstraction of the production lines' indoor environment including a characterization of the devices, facilities, walking ways etc. This helps to understand the complexity and opportunities of the spatio-temporal data of a production line and how indoor GIS can be applied.
Generally, the proposed approach is conceptual at the beginning which results in an empiric approach. The conceptual part is covering the main steps of setting up an indoor environment, creation of a spatio-temporal data concept and extraction of trajectories. The analysis, the spatio-temporal data mining and the comparison of different extracted trajectories are the empiric part of the applied approach.
A conceptional workflow model is provided in figure 4.1 pointing out the four main steps the modeling phase, implementation, data mining and the final visualization. Further it shows the working steps starting with a literature as base work, different modeling steps which lead to the implementation of a spatio-temporal data concept and an indoor network to enable the creation of different transport trajectories. Spatio-temporal data mining supports the analysis of the data through SOM and the visualization of the results which leads finally to a proof of the concept.

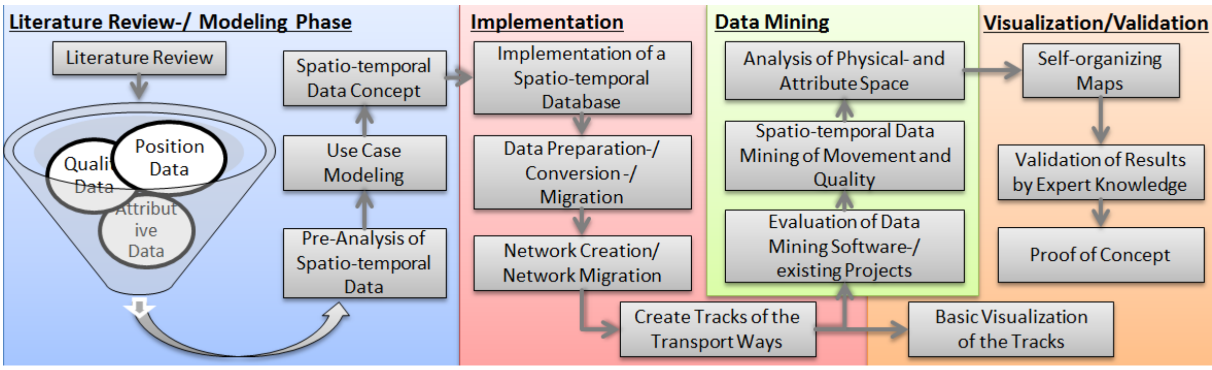


Figure 4.1: Conceptional workflow model of the proposed approach.

4.1 Production Line Under Review

Generally, a production line is a special indoor environment regarding to the layout as it is supposed to support the various production steps in an efficient manner. Therefore, the indoor layout looks different than the layout of an office or buildings for residential use. The production line under review is provided by an Austrian semiconductor company. Due to the fact that the sequence of equipments is changed as well as the layout of the production hall is changed, flexible entities have to be created. Further, the production of semiconductors is done in a cleanroom environment including more constraints. Such a constraint can be the separation of production processes to omit contamination risks or not every production good is allowed to be anywhere in the production line.

To enable and support the transport in the production line indoor environment in a proper way, mainly two entities are existing and interacting. These are on one hand the production goods undergoing various processing steps and on the other hand the production equipment processing the goods. As the semiconductor development is very flexible, also the production environment is changing. Equipment can be reordered in the production line or it is simply replaced by a newer or other equipment. A very important issue is, that the production line under review is not fulfilling the metaphor of a band conveyor or a fixed processing chain as it is commonly expected in the public.

The semiconductor production line under review is a complex and flexible system. This is regarding to the following reasons:

- The overall processing time from the raw material up to the final chip needs from a few days up to a couple of weeks. This varies depending on the product and the priority class.
- A high number of different products in the production line indoor environment have to be processed on different production equipments.
- The number of production steps is up to several hundreds to finish a product.
- Different degrees of completeness per product are present permanently throughout the production line.
- A production step can be carried out on several equipments which are from time to time distributed over different production halls. The processing time and the quality of the production is also depending on the used equipment.

Summarizing, the production line under review is one production hall of an Austrian semiconductor manufacturer. The production line is highly flexible as it does not follow the typical band conveyor metaphor. There is a high number of different productions with a different level of completeness and each product can be processed on several equipments. The choice of the equipment affects the processing time and the quality, but the selection of the equipment is not part of this thesis.

4.2 Indoor Positioning Solution

The used indoor positioning solution has to consider the special production line indoor environment. Therefore, technical drawbacks can occur through the metallic surfaces of for example the equipment, a high number of to be tracked and located production assets and the consideration of the cleanroom restrictions which are essential for the semiconductor industry. The indoor positioning solution combines the different technologies active and passive RFID as well as ultrasound systems to enable the localization of their production assets. This indoor positioning solution is used for a couple of years and proofed itself by demonstrating reliability and efficiency of localization and communication of the moving production assets in the indoor environment *(Thiesse et al., 2006)*.
Requirements for the indoor position solution are including the cleanroom environment, the surfaces of the equipment and the production line and the high number of production assets. Further, the production line is operating *24/7* there is a high variety of different products and a high number of different or similar equipments. The complexity is described in more detail in section 4.1. A special emphasis has to be set on the transportation tasks that are done by human operators between each production step. This is crucial as there are many localization tasks in a small physical area as for example on transportation carts or even on shelves. The localization has also the task to decrease lead time and reduce activities without a certain value *(Thiesse et al., 2006)*.
According to Thiesse et al. (2006) the summarized requirements for the localization are:

- Localization of production assets is greater than 1.000.
- Areas with many shelves can gather more than 100 boxes per 6 square meter.
- Accuracy of < 1 meter to enable a proper search for a production asset.
- Electromagnetic reflections through the metallic production line environment complicate the RFID based localization.
- Localization during a physical movement has to be significantly be < 1 minute.
- Production assets and the RFID receiver are not permanently in a direct line of sight.
- Communicate with the operator to enhance the control of operations.

A combination of RFID and ultrasound is used to be prepared for the accuracy of the localization and to enable the communication. Active RFID is solving the required battery and communication task but not for localization as it was inappropriate through the electromagnetic reflections. The localization itself is computed through ultrasound signals. Therefore, the production assets include active RFID adapted with ultrasound sensors. In this way, a developed algorithm is determining the location via ultrasound and the RFID controller is transmitting the data. Summarizing the RFID is used for an efficient communication, to check the plausibility of the determined location and access operations. Details of the used simplified localization algorithm, the complete system architecture of the localization system and the hardware components are not part of this thesis *(Thiesse et al., 2006)*.

4.3 Description of the Base Data

The analysis of the base data is a crucial issue with respect to data management and setting up a data concept, in this case the spatio-temporal data concept. The base data can be separated into three types of data. These three types are spatial data, temporal data and attributive data. The components can be seen in figure 4.2. The input data is provided in an AutoCAD file as well in several distributed databases.

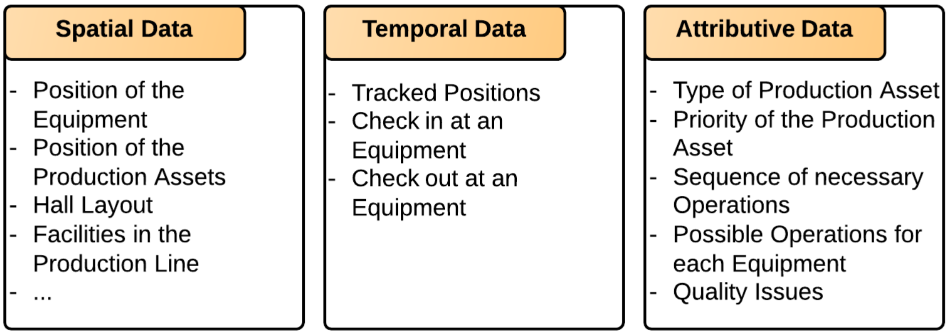


Figure 4.2: Listing of the three types of base data.

Spatial Data:

The spatial data can be separated into data regarding the layout of the production line environment on one hand and on the other hand into spatial data describing the position of the production assets. The spatial data of the production line environment is provided in an AutoCAD file separated in a layer structure. The AutoCAD file covers the equipment in the production line and several facilities like working stations, tables, shelves, elevator and a band conveyor. All the data included in the AutoCAD file need a high effort in the later carried data preparation in section 5.2. Spatial data regarding the production assets is present in a database of the Austrian semiconductor manufacturer. Therefore, the data has to be stored temporarily and prepared for the further analysis. The spatial data is stored as separate coordinates and not as geometry as it is needed. The preparation and migration steps are described in section 5.2 in detail.

Temporal Data:

As it can be seen in figure 4.2 above, temporal data is covering the tracked positions for the production assets and the check in and check out of the production assets for each processing step at an equipment. This temporal data enables possible calculations of for example the overall duration or an approximation of the trajectories by combining of used equipments and tracked positions. The temporal component is always a full timestamp combining the date and the exact time of the event. Temporal data is stored in a database and has to be migrated and prepared for the spatio-temporal data concept that has to be developed. For each event a timestamp is stored if it is either a new tracked position of the production asset, a check in of a production asset at an equipment to start an operation or even the check out of a production asset at an equipment after an operation finished successfully.

Attributive Data:

Attributive data is describing the production assets itself. This takes into account descriptive informations about the production assets as which type it is and what is the sequence of operations that has to be considered. Therefore, attributive data is also covering the equipment which is separated into different groups and their names as well as the link between the equipment and possible operations that can be executed at a specific equipment. Further attributive data concerns the facilities such as tables or types of shelves. Basically, attributive data describes all features interacting in the indoor production line environment as there are equipments, facilities and the production assets. The attributive data is stored in several distributed databases and has to be migrated and prepared for the spatio-temporal data concept. Table 4.1 shows a listing of the attributive data for the production assets, equipments and the facilities. A short description is provided with some given examples of the data. Within this thesis no real values of the Austrian semiconductor manufacturer are provided for security reasons. Therefore, in tables, figures and in the text only anonymized values are used.

Table 4.1: An example of attributive data to get an overview.

Examples of Attributive Data	
Production Asset:	*Priority* - high, low or no priority for the production asset. *Classification* - grouping of the type of production assets. *Route* - sequence of the order of operations. *Start time* - Timestamp 01.01.2014 12:37:45 *End time* - Timestamp 01.01.2014 12:45:45
Equipment:	*Operation* - List of possible operations for each equipment. *Classification* - Grouping of the type of equipment. *Name* - Name for identification .
Facility:	*Type* - Table, working station, shelf, band conveyor .

The base data is providing all necessary information for a possible visualization in a GIS. To gain new knowledge out of the data through spatio-temporal analysis the data has to be pre-processed. The pre-processing phase of the data and the migration in an own developed spatio-temporal data concept is the basis for all proposed tasks.

4.4 Modeling Phase

An important issue is the modeling phase, which is defining the use cases. The use cases are used to prepare the spatio-temporal data concept. This implies the spatio-temporal database which is one of the key issues. These steps can be seen in figure 4.3, where the defined base data are taken into account for the use cases. The use cases are necessary to define the spatio-temporal data concept. The spatio-temporal data concept will combine the spatial data of for example the equipment and the temporal data how it possibly changes over time. This leads later on to a spatio-temporal database which will be described in section 5.1. Other import issues within the modeling phase are the characterization and abstraction of the indoor environment which will lead to an ontology of the production line. This helps to understand the processes and how to navigate in the production line. A special focus is set on the indoor navigation which will be solved through a graph based network approach. Another important issue is dealing with the modeling of the input data for the training of the SOM. This is a crucial part as it is important

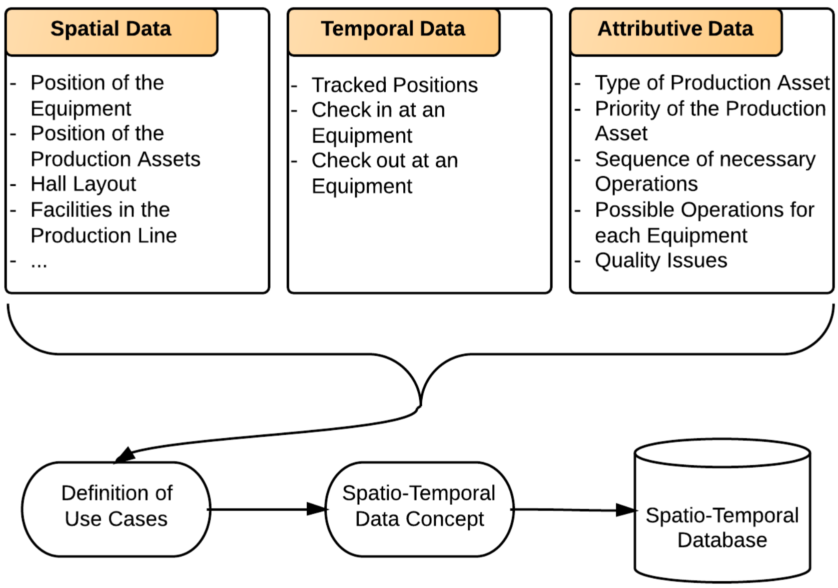


Figure 4.3: Flow model from the input data to the defining of the use cases, the spatio-temporal data concept which leads to the database.

4.4.1 Definition of the Use Cases

A use case diagram is applied to represent the interactions in the system of a production line in a simplified way. It helps to show the different types of actors in the production line system and the different opportunities how the actors can interact. In this use case diagram of a production line system, mainly five actors are identified. These actors are the transport operator, a control operator, the production asset, the equipment and the implemented positioning solution. Figure 4.4 is a use case model and shows the interactions of each defined actor in the production line.

In the following paragraphs each of the five actors is described as well as the interactions between the actors. This simplifies the understanding of the production line and the provided data.
The **transport operator** is probably the key actor within the whole production line system as he is transporting the production assets and is making important decisions about the transportation way and the next equipment. Basically, there are two possible transport opportunities. On one hand a single transportation of a production asset depending on the priority. On the other hand, the transport with a transportation cart. Therefore, a production asset has to be taken which includes the checking of the priority, identifying the next equipment to order the production assets accordingly or the identification of the shelf for a further transportation. If the transport operator has a cart, he has to take several production assets including the same steps as before. Finally is the transport itself.
Control Operator is mainly dealing with data stored in different databases to understand and analyze the production asset. This can be done by analyzing tracking information, used equipments and several information that occurred during the processing or the transporting. The control operator uses the gathered data to provide useful information of for example how to reorder the equipment or new development of recipes. Therefore, this thesis has the aim to show new opportunities to analyze the existing data and to gain new knowledge by the usage of GIS and spatio-temporal data mining techniques.
Another actor in the use case diagram is the **positioning solution** described in section 4.2, having the task to track the production assets and to communicate information to the transportation operator. The stated activities for the indoor positioning solution are the tracking of the production assets which gives an approximation of the overall trajectory and the current or the last position. Therefore, tracking activity includes the storing of the information in the database. The other stated activity is the communication of the positioning solution. This is regarding to the display of the tracking devices highlighting information for the transport operator in the production line. Enabling the display of priorities, operations and further information.
Further, the **equipment** provides information about the processing of a production asset. The activity of the equipment is clearly the processing of the various production assets. The processing itself includes the storing of information in the database system. This information is including an identifier of the production asset and quality parameters and useful information for an analysis later on. The location can be considered as the equipment itself, which can be changed over time. This is not included in the use case diagram in figure 4.4 as it is not part of the processing and transportation tasks.
Finally, the **production asset** itself which is another key issue in the production line. The production asset has mainly two activities which occur before and after the production asset is processed at an equipment. First, it is checking in to be identified at the equipment. Second, after the processing at an equipment itself

the production asset is checking out. All these data includes the storing of the information in the database, to provide the control operator in the end more information that can be analyzed.
A detailed listing of each use case can be seen in appendix A. The listing includes information about the initiating actor, the goal of the actor, the pre-condition and the post condition of the use case.

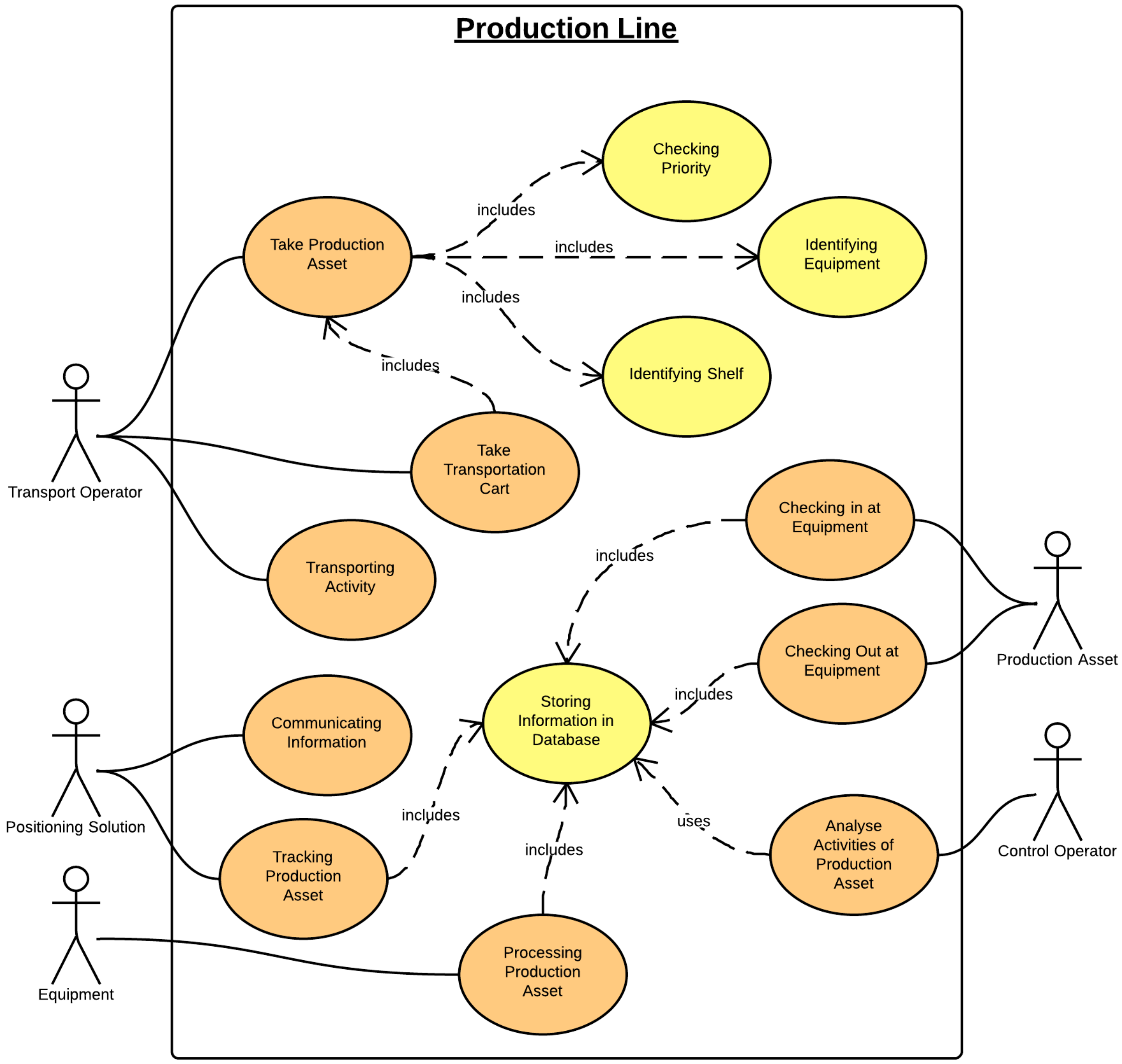


Figure 4.4: Use case diagram representing interactions of actors in the production line system.

A basic flow of the production line system looks like the following. The transport operator is taking either one production asset and is checking the priority, identifies a suitable equipment or a shelf as target location for the production asset or a transportation cart where he has to take several production assets under the same preconditions which leads to the transportation task. If the production asset arrives at an equipment it is checking in, the equipment processes the production asset until it is checking out. Meanwhile, the information is stored in a database. The whole process is tracked through the positioning solution which is transmitting the position to the database and useful information for the transportation is communicated. Overall, a control operator has the ability to access the data stored in the database and analyze the activities of production assets.

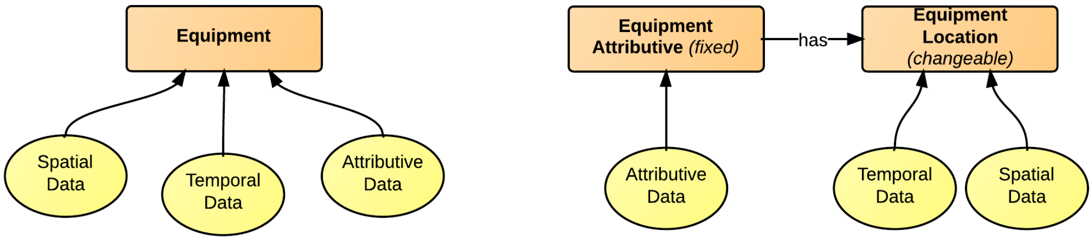
4.4.2 Spatio-Temporal Data Concept

The spatio-temporal data concept is the basis for the later on development of a spatio-temporal database. It is including the use cases from sub-section 4.4.1 that has to be applied on the base data from section 4.3. The spatio-temporal data concept has to provide the ability to answer different questions about the production asset, equipment and facilities. The challenge is to provide a data model to answer all questions that were stated during the use case modeling. These question are concerning either temporal or spatial aspects or even combination of both. Some example questions are listed in table 4.2.

Table 4.2: Example questions that have to be supported by the spatio-temporal data concept in an effective way.

Examples of possible Questions	
Temporal Questions:	When was a movement? How long was a product processed? How long was the transportation? How long was a product standing? ...
Spatial Questions:	Where was a production asset? Where is the next transportation target? What is the transport way? In which room was the production asset? ...
Spatio-Temporal Questions:	When was the product in the room? Where was the product at 09:00 a.m.? Which products where in which room at a certain time? ...

To answer these and similar questions, the attributive data, the spatial data and the temporal data has to be ordered and has to be brought in a specific structure. This is very important for the production assets, the equipment and the facilities in the indoor production line environment. The changeable location of the equipment, facilities and the production assets have to be considered as well as the temporal connection between production assets and equipment or facilities. Figure 4.5 shows how the concept can look like in a simple way according to the equipment. Therefore sub-figure 4.5a, identifies the equipment itself in the production line with the three types of input data *(spatial-/, attributive-/ and temporal data)*. In the spatio-temporal data concept in sub-figure 4.5b there is a main equipment object called equipment attributive, which is including all the general attributive information of an equipment that are more or less fixed. Additionally, each equipment object has the equipment location covering the temporal and the spatial data. This means that each equipment has one location for a certain time and that the location is changeable over time.



(a) Equipment with the three types of base data as input.

(b) A main equipment object, has a changing location over time.

Figure 4.5: An example of the spatio-temporal data concept based on the equipment in the production line.

To model the spatio-temporal data concept in an abstract way an entity-relationship model is used. It describes the data including the requirements and process requirements in a simplified way implying the connections between the data. Figure 4.6 shows the complete entity-relationship model. The spatio-temporal component is included using the method described for the equipment in figure 4.5.
As it can be seen in figure 4.6, the **production asset** is the main component on top of the figure including the attributes for the representation, priority, timestamp from and to as well as a unique identifier. By considering the equipment example from before, it can easily be seen that each equipment has one or many positions including the geometry, a timestamp and the unique identifier. The unique identifier provides the capability to connect the attributive data with the position. Further, the changing positions or in other word the movement can be identified by ordering the positions by the timestamp.
The same method is applied on the **equipment** and the **facility**. Each equipment has attributes like the name, a classification, a description and an ID as well as the changing position with the id, geometry and timestamp from and timestamp to. The equipment has at least one or many positions and each position has one equipment. Exactly the same has to be done with the facilities, except there are other attributes for the facility itself and the position attributes have to be the same. A connection between equipment and facility is kept by storing the geometry of a **room**. Through the spatial connection it can be defined that each equipment has to stand in one room as well as the facility has to stand in one room. In the other side, each room has no equipment if there are only facilities or many equipments and vice versa.
To store the connection between the production asset and the equipment another

entity is called **Production Asset in Equipment**. This entity has an ID of the production asset, an ID of the equipment and the check in and check out of an production asset by the equipment. Each production asset or equipment has no connection or many connections and each production asset in equipment entity has exactly one equipment and one production asset. Another entity is **Production Asset in Facility** having the same features for the production asset and the facility storing information if a production asset is standing in a facility.

The last two entities are **Recipe** and **Operation**. Therefore, each production asset has one recipe and a recipe has no or many production assets. The attributes of the recipe are an ID, a description, and an operation including the next operation. This shows that each Recipe has many operations as well as each operation has many recipes, in another sequence. Therefore, a specific recipe has a unique order of operations. To close the connection between the production asset, recipe, operation and equipment, each equipment has several operations and an operation can be executed on several equipments.

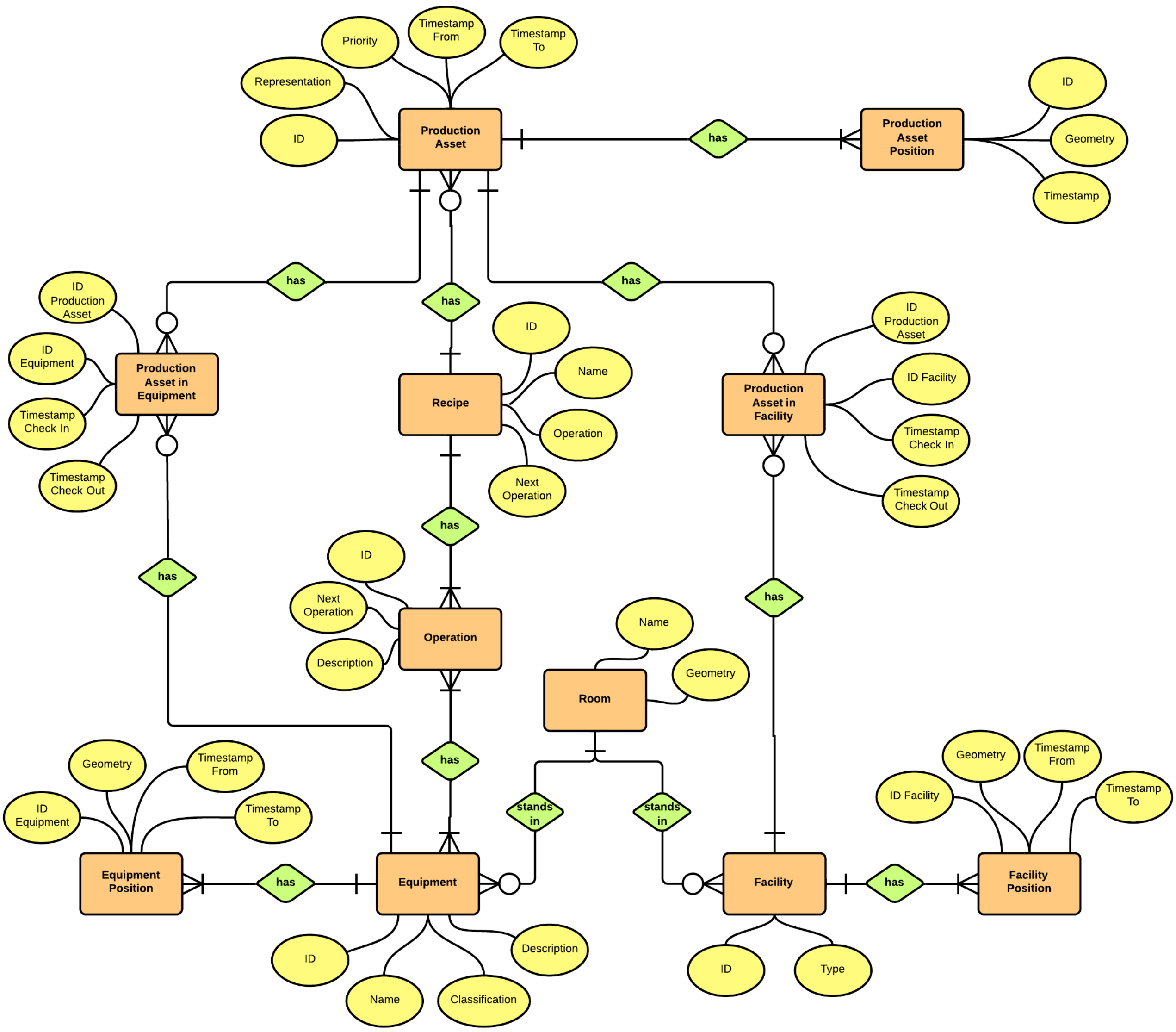


Figure 4.6: Entity-Relationship Model illustrating the spatio-temporal data concept in an abstract way.

4.4.3 Characterization and Abstraction of the Indoor environment in a Production Line for an Ontology

The complex and changeable production line indoor environment has to be characterized and abstracted to understand the behavior of the production line. The indoor environment is supposed to support on one hand the routing in the indoor environment where some approaches are described in section 3.3. On the other hand it is important to show the production line features such as the equipment in an abstract way with all their changes and types of intersection. Partly some approaches for the indoor environment are described in section 3.4. Generally, there are the two important types for the routing and for the indoor environment.
The indoor environment includes possible walking paths, the processing units or equipments, the facilities and the production assets. Further the indoor environment has also to include specific barriers. Table 4.3 shows the identified features for the indoor environment.
The routing in the indoor environment is based on a graph structure consisting of edges and nodes. For the routing itself, nodes have to be assigned to the equipment and the facilities to include them in the routing. As well as there must be an accessibility to enter the indoor production line environment and to exit it. Table 4.4 is listing the identified features for the routing which are mainly nodes and edges. As it can be seen in table 4.3 and 4.4, routing and indoor environment characteristics have overlapping features with different tasks or restrictions.

Table 4.3: Defined features of the production line environment.

Characterization of the indoor production line environment	
Equipment:	– Has a specific shape in the production line. – Has operations that have to be executed. – Changing position over time.
Facility:	– Has a specific shape. – Type of facility. – Fixed or changing position.
Production Asset:	– Occurs in a Facility, Equipment or it is transported. – Changing position over time.
Hall Layout:	– Specifies rooms and corridors – Walls are either fixed or changeable *(Outer wall or inner wall)*. – Different floors and different accessibility through elevators or stairs.
Barriers	– Occur through the hall layout, equipment or facilities *(Stairs, walls,...)*. – Security issues and restrictions for operators. – Quality issues as for example air quality.

Table 4.4: Features for the routing in the production line.

Characterization of features for the routing indoor	
Production Asset:	– the product that has to be transported or navigated in the production line.
Edges:	– Possible walking ways between equipments. – Possible walking ways between facilities. – Corridors in the hall layout.
Nodes:	– Each equipment in the production line. – Each facility where the routing is useful. – Navigation nodes in corridors and rooms. – Nodes to enter the production line environment. – Nodes to exit the production line environment. – Nodes to transfer between floors. – Barriers as stop or end nodes.

4.4.4 Data Modeling for the Self-Organizing Map

The modeling of the data for the SOM is an important issue. First of all, the general workflow of a production asset is described, to understand the behavior in an abstract point of view. Further, different opportunities to prepare the data for the training process of a SOM are modeled. This is necessary, to understand the SOM itself and to be able to interpret the results of the SOM and compare it with the data. Basically, two types of a SOM are created for the production assets. On one hand a SOM is including the production assets itself and the used equipment. On the other hand, another SOM is describing the quality issues and as well the used equipment. Therefore, an adequate data model has to be developed.
In figure 4.7, the general workflow of the production asset moving through the production line can be seen. Each production assets provides a sequence of equipment that has to be used or that was used. In this example, production asset I has an overall start and end time of all processes in the production line including a sequence of equipment. The equipment is visualized on the timeline for the production asset according to a check in and check out of the production asset at the equipment. The processing of the production asset at the equipment is stored through timestamps and the processing time can vary, which is not visualized in the abstract model. Production Asset II has a different sequence of equipment and a different time range of the overall start and end time that can be seen by the appearance of the first equipment. This timeline is possible for each of the production assets according to the production asset *n* having another sequence of used equipment stored in the spatio-temporal database.

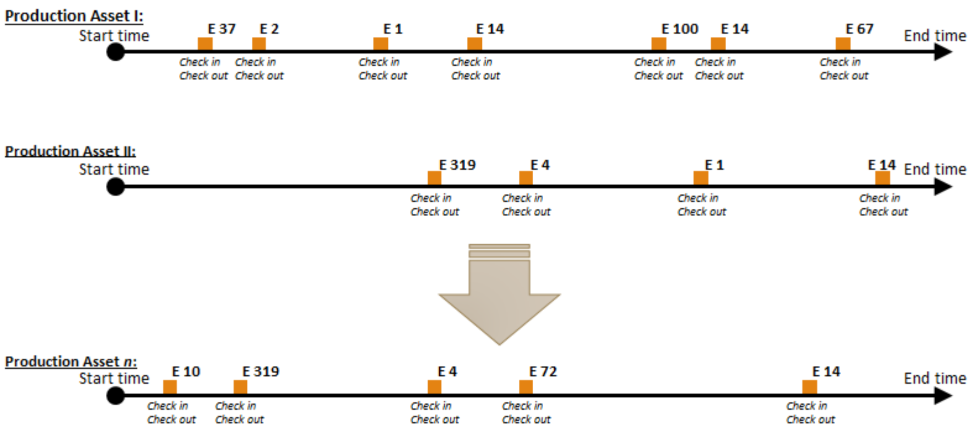


Figure 4.7: Abstract model of the production asset moving through a sequence of equipment over time.

Three different examples as data basis for the training process of the SOM are taken into account. The first approach is basically storing the sequence of the used equipment in a table. The second approach considers the frequency of the equipment per production asset and the third approaches summarizes the sequence of the equipment for each production asset.

Sequence of Equipment Data Model

The first approach is creating a matrix of production assets and equipment. Therefore, the sequence of the appearance of equipment is stored to compare the overall sequence and to find similarities. Similarities can be identified by a normalization for each column. Missing values can occur if equipment are used more often, which is a crucial issue for the described production line environment. Figure 4.8 highlights the possible data structure for the training process. Further, the timeline of the production assets from figure 4.7 are used within this example. As it can be seen, production asset I uses equipment 14 two times, highlighted within the red rectangle. In the table it cannot be seen, as only atomic values are considered in the database. This issue can lead to wrong or misleading interpretations.

	E1	E 2	E 4	E 10	E 14	E 37	E 67	E 72	E 100	E 319
Production Asset I	3	2	0	0	4	1	6	0	5	0
Production Asset II	3	0	2	0	4	0	0	0	0	1
Production Asset *n*	0	0	3	1	5	0	0	4	0	2

Figure 4.8: Model of storing the first appearance of an equipment.

Frequency of Equipment Data Model

In comparison of the first approach, the second data model is storing how often the equipment is used. A matrix is created of production assets and equipment. Additionally, for each production asset it is counted how often one equipment is used, as it is known that a production asset can be processed by the same equipment several times. This covers questions concerning the overall similarity of sequences of production assets. The following figure 4.9 shows the resulting table of this approach. Random test values are used to show either a similarity between the sequence of equipment or not. As it can be seen production asset I and Production asset n are in this case more similar than production asset I and production asset II. This can also be seen through the green or red connecting line.

more similar (Production Asset I – Production Asset *n*)
less similar (Production Asset I – Production Asset II)

	E1	E 2	E 4	E 10	E 14	E 37	E 67	E 72	E 100	E 319
Production Asset I	12	0	0	37	4	0	2	1	15	21
Production Asset II	0	7	19	2	0	34	15	4	0	3
Production Asset *n*	15	1	0	34	8	1	0	0	17	24

Figure 4.9: Model of storing the frequency of the equipment.

Timely order per Production Asset Data Model

The third data model is creating a matrix of equipment and production assets. The columns are the production assets and each row is one production process at one equipment. Sequence is provided as each process is stored in the timely order of all processes. This is a Boolean approach, where the row describes mainly the equipment with the check in and check out of the production asset as well as an identifier of the production asset. Each production asset has an on column and the Boolean approach identifies the corresponding production asset. By adding the identifier, it is possible to later on project further information onto the SOM.
Figure 4.10 shows an example of this proposed approach relating to the sequence timelines of the three production assets in figure 4.7. P.A. stands for an abbreviation of production asset and the timestamps IN and OUT represent the check in and check out of a production asset at the equipment. The Boolean approach can easily be seen as well as the equipment is in the timely order. This is a resource intensive approach, as there is a high number of production assets which defines the number of dimensions of a SOM.

	P.A. I	P.A. II	P.A. *n*	*IN*	*OUT*	ID
E 10	0	0	1	12:00	12:10	n
E 37	1	0	0	12:23	12:32	1
E 319	0	0	1	12:37	12:44	n
E 2	1	0	0	12:46	12:49	1
E 1	1	0	0	12:53	13:02	1
E 4	0	0	1	13:07	14:32	n
E 319	0	1	0	13:07	14:32	2
E 14	1	0	0	15:07	15:15	1
E 4	0	1	0	17:56	18:04	2
E 72	0	0	1	17:56	18:04	n
E 100	1	0	0	18:15	18:24	1
E 1	0	1	0	18:33	18:59	2
E 14	1	0	0	18:59	19:07	1
E 14	0	0	1	19:28	19:44	n
E 67	1	0	0	20:02	20:14	1
E 14	0	1	0	20:39	20:55	2

Figure 4.10: Model of storing the order of the equipment per production asset.

4.4.5 Definition of a Quality Issue affecting the Production Asset

As it is described in the previous section 4.4.4, each production asset has a specific sequence of used equipment through the whole production line. Quality measures can be encountered within such a sequence by the equipment and can have an effect on the production asset. Spatio-temporal data mining has the ability to analyze such issues. Therefore, quality issues have to be defined and an appropriate data model has to be applied. Possible three data models are already described in section 4.4.4 including examples for the overall sequence of a production asset.

Defining the type of quality issue is similarly to the data modeling for the SOM in section 4.4.4 an important issue to understand the results and to be able to interpret them accordingly. In general, within this research three different types of quality issues are considered within this research. First of all, an evolving quality issue is considered by a sub-sequence of equipment per production asset. Thereby, the sub-sequence is created considering all the equipment before a quality issue or in between two quality issues. Second is a pre-defined number of equipment that has to be considered for a quality issue, which sets also up a sub-sequence with equal length. The equal length of the sub-sequence is limited through a threshold of for example three or four equipment that set up a quality issue sub-sequence. The third approach includes only the triggering equipment, so basically the last equipment in the sub-sequence. The last approach is mentioned separately even if it is similar than the second, as another type of analysis has to be carried out.

Quality Issue defined as a Sub-Sequence of Equipment

Generally, a quality issue is occurring after the check out at the equipment. Thus, it involves the equipment as well as the production asset. Triggered is a quality issue by the tracked production asset. The first definition of a quality issue is dividing the overall sequence of equipment per production asset in a subset of sequences with different length. The different length is resulting by the appearance of the quality issue in the overall if it is the first, the last or a equipment in between. If a production asset has for example 3 triggered quality issues overall, there are three different sub-sequence. In this case, a sub-sequence is defined to be a sequence of each equipment until the quality issue is triggered. If there has already been one quality issue, the sub-sequence includes each equipment from the last quality issue until the next triggered issue. The last equipment in the sub-sequence is triggering the quality issue.

In figure 4.11 examples are given for the overall sequence of a production asset without a quality issue and a production asset with quality issues defining the sub-sequences. Another example shows how the sub-sequence of different production assets can be compared. In sub-figure 4.11a, the overall sequence of a production asset is given. The overall sequence is *"E 37-E 2-E 1-E 14-E 100-E 14-E 67"*, whereas the character E is the abbreviation for equipment.

In comparison, sub-figure 4.11b shows an example of the same production asset with 2 quality issues triggered after an equipment. Therefore, the overall sequence is split up into the two following sub-sequences. The first quality issue creates the sub-sequence *"E 37-E 2-E 1-E 14"*. The sub-sequence of the second quality issue starts after the last quality issue which is in this example quality issue 1 and stops after the event of the quality issue. The second sub-sequence is *"E 100-E 14-E 67"*.

Sub-figure 4.11e states the sequence of two different production assets having a specific sequence of equipment in common until a quality issue occurs. The quality issue does not have to be at the same timestamp, but in the same order of the overall sequence. As it can be seen in the example production asset III has a sub-sequence of *"E 2-E 17-E 191-E 4-E 7"*. In addition, production asset VIII has a sub-sequence of *"E 88-E 293-E 401-E 4-E 7"* until the quality issue. On the basis of this example, equipment *"E 4-E 7"* are in the same order in the sub-sequence of both production assets and shows a certain similarity before the quality issue occurs.

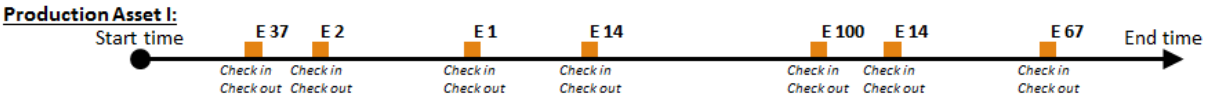
Quality Issue defined as a Sub-Sequence of a fixed Number of Equipment

The second type of quality issue is created by a pre-defined and fixed number of equipment considered for the sub-sequence which can be used for an analysis. This limitation of equipment has to be handled carefully and mainly bases on a decision of a manufacturing site providing the data of the production assets. Such a quality issue can be for example the thickness of a production asset which is checked in regular steps. This threshold for the creation of the sub-sequences depends on the general quality of a production asset, the value of a production asset and the likelihood of quality issues that can occur. An example can be, that a manufacturing site defines a threshold of four equipments per such a quality issue sub-sequence covers always the same length of equipments, except if it is one of the first three equipments in the overall sequence. In order to set the focus on the previously used equipment, even if the quality issue occurs after a higher number of equipments. This type of quality issue ensures a comparison of the equipment with a high likelihood on having an influence on the actual quality issue as the last four equipments are used.
In figure 4.11 examples are given for the overall sequence and the different defined types of quality issues. In addition to sub-figure 4.11a where the overall sequence is visualized and sub-figure 4.11b covering the first approach, sub-figure 4.11c shows an example with a threshold of two equipments of setting up a sub-sequence of a quality issue. Thereby, the threshold means in this case that the company defines that a quality issue includes only the last two equipments as they have several ongoing checks verifying the quality regularly. As already mentioned before, this threshold can also be higher. The overall sequence can be seen in sub-figure 4.11a and is defined to be *"E 37-E 2-E 1-E 14-E 100-E 14-E 67"*. The sub-sequences created by a threshold of two equipemnts per sub-sequence can be seen in sub-figure 4.11c. According to this example, the sub-sequences are as follows *"E 1-E 14" "E 14-E 67"*. This example shows that *E 14* is involved in both sub-sequences, which can be interpreted that *E 14* is likely to evolve a quality issue.

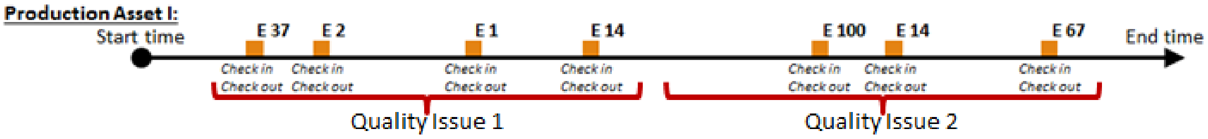
Quality Issue defined by the triggering Equipment

The third type of quality issue covers only the triggering equipment. This means in detail, that the quality is checked after each processing step at an equipment. Regarding to this definition, the equipment triggering the quality is directly recorded. Depending on the type of quality issue, which can either be a measurement value or for example the proof of an order of objects on a production asset, this approach is useful. If the order of objects on a production asset essential and there is a change at an equipment this is an appropriate approach. Under the consideration of a measurement value, a measurement value has a certain threshold where a quality issues is triggered. Therefore, the disadvantage of the current approach is that only the last equipment is considered where the threshold is reached, whereas a big step to this threshold occurred an earlier equipment.
Again in figure 4.11 examples are given for the overall sequence and the different

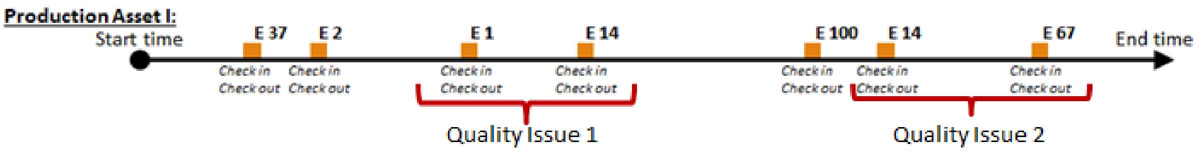
defined types of quality issues. Sub-figure 4.11d gives an example for the currently described approach. The equipment causing a quality issue is for quality issue 1 *E-14* and for quality issue 2 *E-67*. By investigating quality issue 2 it can be seen that *E-14* is directly before the triggering *E-67* and *E-14* is already triggering quality issue 1. Depending on the type of quality issue, for example measurement value or order of objects, this knowledge may be important.



(a) Overall sequence of equipment of one production asset without quality issues.

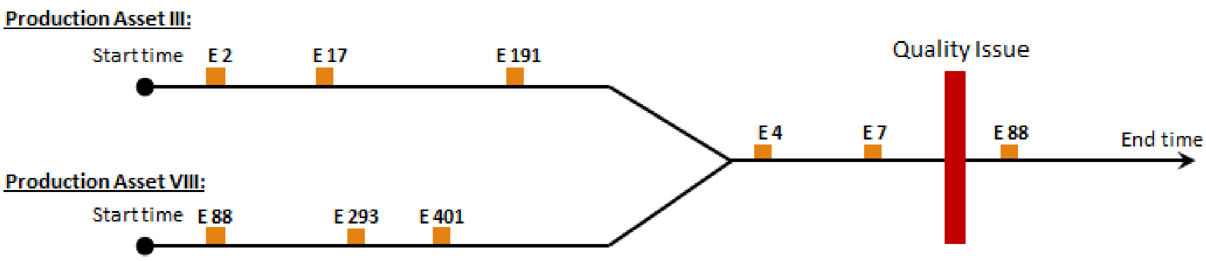


(b) First Approach - A production asset with two quality issues and how the sub-sequences are defined.



(c) Second Approach - A production asset with two quality issues and how the sub-sequences are defined.

(d) Third Approach - A production asset with two quality issues and how the sub-sequences are defined.



(e) Different production assets resulting in a quality issue with some equipment in common.

Figure 4.11: Definition examples of quality issues and their sequences.

Chapter 5

Implementation

The implementation chapter covers the spatio-temporal database including the spatio-temporal data concept as well as further data such as the network for the production line. The data preparation and migration executed to set up and to define the whole base data for the analysis. Another part of this implementation chapter is the migration of different distributed databases and to bring the data into the right format. This are crucial issues to have a proper result in the end of the analysis. Further, the creation of a network to enable the routing in the production line is an essential issue. This includes the features for the routing in the production line in table 4.4. An ontology of the indoor environment is created which describes the whole characterization and abstraction of the indoor production line environment based on a graph and the transporting of production assets. A general issue is also the extraction of movement paths in the production line environment. Therefore, different paths are possible such as the approximated path of the production asset or the shortest path through the overall processes.
To analyze the data, data mining is applied within this section. The data mining is regarding to the analysis of the movement behavior of similar or non similar objects and movement paths. Self-organizing maps are created of the selected data in section4.4.4. Component planes will be analyzed and compared as well as there is an analysis of the physical space and the attribute space including the time.

5.1 Spatio-Temporal Database

As already mentioned in the previous chapters, the spatio-temporal database is one general data pool. To include all the necessary features for the routing in the indoor environment as well as the spatial and temporal concept, the open-source object-relational database management system PostgreSQL is used with two extensions. The two extensions are on one hand the PostGIS extension to enhance the database management system with the spatial component and on the other hand pgRouting, an open-source routing library for PostgreSQL/PostGIS. The database management system and the extensions are selected as one advantage is the open-source availability as well as the provided features and capabilities of the extensions. The used components and their tasks can be seen in figure 5.1. The main point is the PostgreSQL database management system including the extensions for the spatial component and the routing ability as well as the temporal component. For the temporal component a database model based on time-stamping which is theoretically described in sub-section 3.2.1 of this thesis.

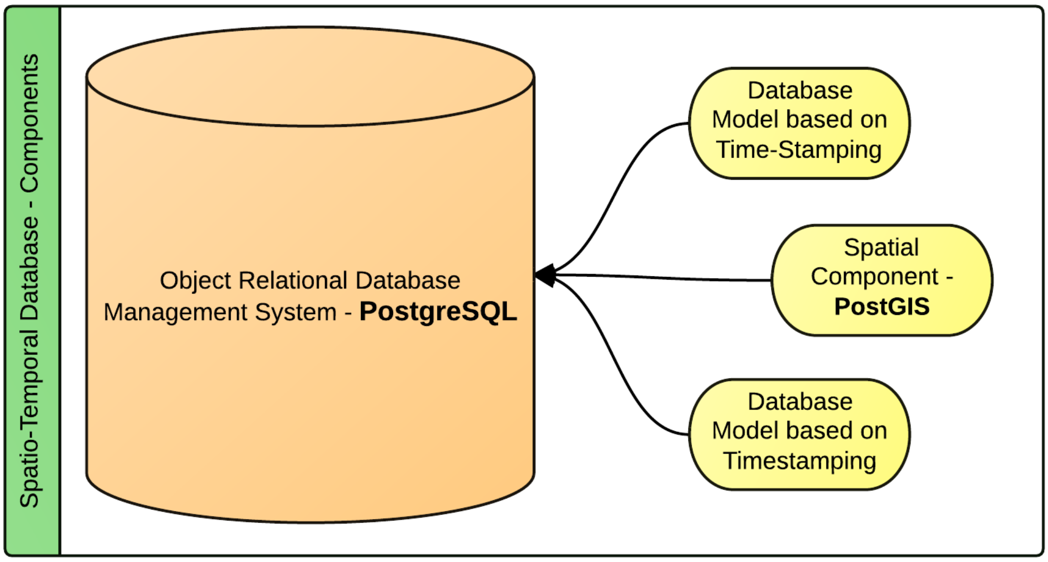


Figure 5.1: Components used to set up the spatio-temporal data concept.

To sum up the selected database model based on time-stamping from sub-section 3.2.1 it is tracking the state of an object which consists mainly of a tuple of timestamps. Therefore, a start and an end time or a state for the end have to be defined. Further, the developed spatio-temporal data concept from sub-section 4.4.2 shows already that the attributive data has to be kept separate from the location and the time component.
Another important aspect that has to be considered within the database is the network which consists of edges and nodes. The network itself is created using pgRouting within the database and importing a predefined network from a shapefile format. The creation of the network is described in detail in section 5.3. The database model has to be prepared to store the data, especially the nodes and the edges. As the characterization of the indoor production line environment investigated that there are different nodes for the equipment, facilities and several access or transfer nodes they have to be considered in the spatio-temporal database. Therefore, three types of nodes are defined. These are equipment nodes, facility nodes and access nodes with a type to distinguish between a transfer or an access. A snapshot of the created relational database model can be seen in figure 5.2. The relational database model is created with the free version of the SQL Power Architect, which provides an easy functionality and has the capability to export the created model direct into a PostgreSQL database. The geometry is not sup-

ported in this version and is in the model replaced with a placeholder attribute called the_geom of type other. Figure 5.2 highlights in the middle the equipment itself with different describing attributes, classifications and an integer value as unique identifier. On the left side the connection with the temporal component of the equipment. Each equipment is only once in the attributive table, but it can have several locations changing over time. Therefore, the unique identifier has to consist of the unique identifier of the equipment and a combination of the starting time. To establish a possible connection for the network, each equipment has another connection to exactly one node representing the equipment in the network.

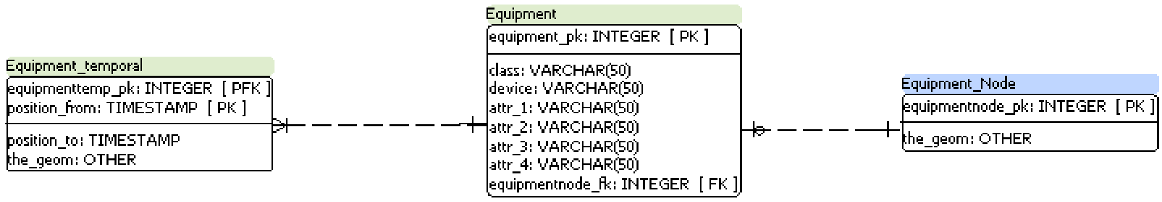


Figure 5.2: Example of the spatio-temporal database based on the equipment.

Summarizing, the implementation of the spatio-temporal data concept is done in PostgreSQL with the PostGIS extension and pgRouting extension. A relational model is created and applied within the database management system. For the production assets, the facilities and the equipment the database model based on time-stamping is applied. Three types of nodes are established to enhance the connectivity to a network that has to be created. The spatio-temporal database includes the described connections and possibilities from figure 4.6, tables for the hall layout and further layers for visualization purposes and important is the network that has to be included by edges and nodes.

5.2 Data Preparation and Migration

The data preparation as well as the data migration is implemented using the Feature Manipulation Engine *(FME)* and different JAVA applications to fit the spatio-temporal data into the developed spatio-temporal data concept. Two different steps of data preparation are necessary. First is the preparation of the AutoCAD input file described in section 4.3. Therefore, the AutoCAD file is converted with FME and prepared with ArcGIS. Second is the conversion and preparation of the base data stored in distributed databases. This data is either converted using developed JAVA applications or also the FME workbench. The general workflow of the data preparation and data migration can be seen in figure 5.3.

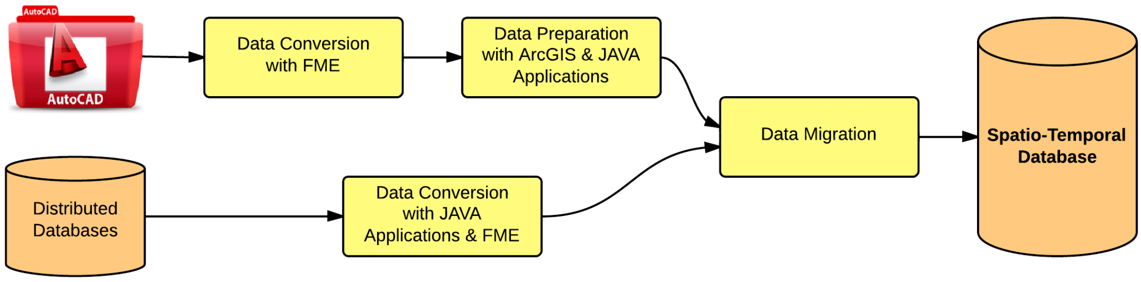


Figure 5.3: Workflow for the data preparation.

Data Preparation and Migration of AutoCAD input data

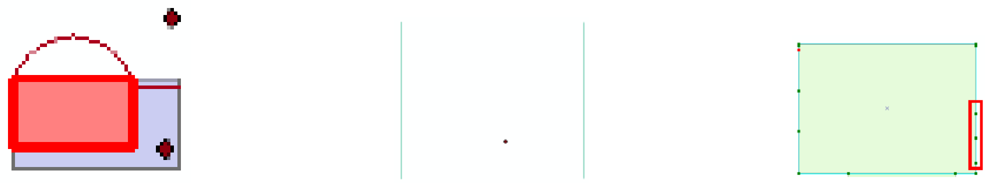
The first step of the data preparation starts with the AutoCAD input data. Basically, the AutoCAD data is converted using the FME Workbench. FME converts one layer of the AutoCAD file into several Shapefiles regarding to the geometry. In most cases, there are four layers covering polygons, lines, points and points storing the textual information provided in the AutoCAD file. At this point, there are maximum four Shapefiles for one type of feature. The data preparation and the merging of files and attributes was executed in ArcGIS after clipping the files to the given study area of the production line. The data preparation starts with a definition of the used geometry, correction of the topology such as overlaying features, simplifying the geometries and the adaption creation of attributes stored as textual information in one of the point Shapefiles provided after the FME converting of the input file. In some cases, also a JAVA Application is used to split up the spatio-temporal information into the pre-defined data concept from section 4.4.2.
One example of the data preparation of the converted AutoCAD files is represented by taking the example of the production line layout. In this case, four different layers representing polygons, lines and two point layer have to be prepared carefully. The polygon layer has to be split up as some lines were connected during the converting using FME and the layout of the production line is defined as a polyline layer. Therefore, new lines have to be added manually, polygons have to be deleted or even split up into lines with the help of the *Polygon to Line* data management tool of ArcGIS. The point layers are not considered for the preparation as either the textual information or the point itself was not relevant for the further processing and visualization of the polyline layer of the layout. To differentiate between the different layout types of the wall and doors attributes are defined. The attributes are listed in table 5.1. These attributes are necessary either for visualization purposes or for further task where it is important to know the type of door. The Shapefile is loaded into the before created database using the Shapefile to PostGIS loader *shp2pgsql*.

Table 5.1: Attributes for the production line layout shapefile.

Defined attributes for the production line layout and doors	
Door:	A standard door within the production line. It can be a main entrance, or any other door in the building
Emergency Exit:	Emergency Exits are extracted and visualized separately for possible future tasks.
Sliding Door:	Sliding Doors are often at the entrance of a room in the production line. They are as well often between the buildings.
Wall:	This category includes all walls that are no outer wall of the building. It is also no wall between buildings or a corridor connecting buildings.
Outer Wall:	The outer Wall of the building as well as corridor walls connecting different production halls.

Another example of the data preparation is considering the equipment and converting into the spatio-temporal data concept. The specific AutoCAD layer is also converted to Shapefiles resulting in the four layers of type polygon, line and 2 point layers. Similarly to the production line layout example from before, also for the equipment the polygons are prepared that they are topologically correct including no overlaying objects. Further, some equipments are represented as lines and they are converted to polygon features and merged to one layer together with the polygon. The points for each equipment are shifted to the edge of the equipment in the inner side of the corridor and will be further used for the network as equipment node. Through spatial joins the textual information stored in the second point Shapefile after the converting with FME are adapted and kept. After these steps the attributes are adapted as it was defined before and the geometries are topologically correct. In the next step, the Shapefile to PostGIS loader *shp2pgsql* is used to convert the Shapefile to a SQL-file format. Before this file is inserted into the database, a JAVA application is used to split up this file into one file storing the attributive information and one file storing the geometry and the provided timestamps. These new SQL-files of the equipment fit to the defined format as it can be seen in figure 5.2 and can be inserted into the spatio-temporal database.
The data preparation itself is a crucial issue for a further accurate analysis. Manually effort is necessary to re-digitize features or even create new features. Examples why the data preparation is important can be seen in figure 5.4. In figure 5.4a, one equipment is converted with FME into the four files. As it can be seen the equipment should be a polygon but there are also line features. These line features have to be either converted to polygons or removed. There are as well overlaying polygons that are highlighted through the red color and two points can be seen. One point is representing the equipment and one point is storing the textual information. After the preparation there is only one polygon and one point which will represent the equipment later on in the network. Figure 5.4b shows a table after the converting process existing of two parallel lines and only one point. The point

stores the textual information that this feature is a table and two further lines have to be added to close it to a polygon. This special case can occur as AutoCAD is not storing topology and on both sites with out a line other features have been in the base file in another layer for example a wall or a equipment. Figure 5.4c shows a polygon which is drawn with too many points after the converting. Some points are highlighted in the red rectangle.



(a) Converted Shapefile before preparation.

(b) Converted table with two parallel edges.

(c) Converted polygon with several redundant points.

Figure 5.4: Examples to highlight the importance of the data preparation.

Data Preparation and Migration of distributed Databases

The second step of the data preparation deals with the inserting of on one hand attributive data from the different distributed databases. On the other hand with position data of the tracked production assets. For this data preparation step again the FME workbench is used as well as developed JAVA applications. The attributive data is migrated directly into the spatio-temporal database. FME modules enable the direct migration by using tools like the *SQLCreator* and the *SQLExecutor*. The SQLCreator provides the opportunity to load data from a database via a SQL query that has to be defined into FME. By connecting this SQLCreator with the other tool the SQLExecutor it is possible to add further queries from another table or another database to gain the data in the right format. The SQLCreator is mainly used to query the relevant data from the distributed databases and the SQLExecutor provides the opportunity to compare it with the data stored from the Shapefiles in the spatio-temporal database. This enables a further comparison such as for example, the already inserted equipment in the spatio-temporal database has an own identification number. This identification is used to obtain a semi-automatic method to extract the data from distributed databases. A number of different developed FME models is used to insert the attributive data for the operations, the production assets, the recipes and the information about the connection when a production asset was in an equipment or a facility accordingly to the spatio-temporal data concept. As it can be seen in the FME model in figure 5.5, the SQLCreator identifies all data that is important and the two SQLExecutor identify the needed IDs in the spatio-temporal database. After all attributes are identified the matched entries are inserted into the table of the spatio-temporal database. The example inserts data for the time of a production asset in an equipment.
Information about the positions of the production assets are stored as well in one

distributed database, whereas the position is not stored as geometry it is stored as *X* and *Y* coordinate in separate columns. Therefore, the data is migrated into the spatio-temporal database in a temporary table. In the next step, the information about the production assets position and the timestamps are stored in a SQL-file for further processing. The processing is done in a JAVA application the same way as it is described with the equipment example before. The JAVA application is generating a new SQL-file with the statement to enter the point geometry by combining the *X* and *Y* coordinate with the PostGIS statement *ST_GeomFromText(Point(X, Y))* which allows the inserting of the data in the spatio-temporal database.

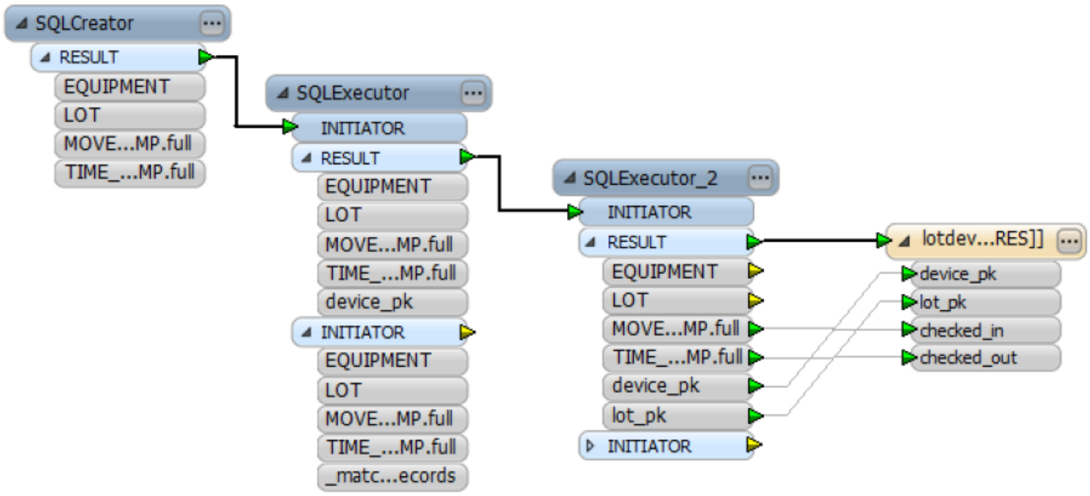


Figure 5.5: One FME module to insert data in the spatio-temporal database.

5.3 Network Creation for the Production Line

To enable the creation of paths and to route within the production line a network is necessary. The network is obtained using ArcGIS as well as FME to obtain features automatically and adjust attributes. As it is mentioned in the equipment example in section 5.2, for each equipment a point is placed inside the corridor or the room. This is necessary to enhance the equipment with the routing ability in the network, which is essentially for the routing from equipment to equipment. Points are also placed on the inner side of facilities that have to be included in the network, such as shelves where production assets have to be transported or where they are waiting for the transport. These shelves are identified by inspection in the production line. Further, access nodes are defined which are described in the characterization of the production line in section 4.4.3, where nodes are necessary to enter or exit the indoor production line environment as well as transfers between for example floors. Access nodes are also used To close the network and enable the transfer access points are added to the network. This is important to have a closed network. In future, these access points can also be the connection between outdoor geography and indoor geography. The network points or nodes are added manually to the border of each facility and each equipment as well as the access nodes.

The access points are mainly used for entrance, exit or a transfer between a hall and even a floor. That means the access point has a general type, access type and the floor. The access type is used to provide the difference if it is either an entrance, exit or transfer. Transfer is generally the movement along a corridor, a stair or an elevator between different floors or halls. The type is storing the information if it is a door, an elevator, a stair or a corridor. To enable the transfer between the different floors also the floor is stored as a number. That means the elevator is linking the first with the second floor, which is important for the network. This means there are four types the elevator, door, stair or corridor and three transfer types such as the transfer, entrance and exit.
Next to the nodes for the network, the edges have to be created with the help of the FME Workbench. Therefore, the corridors and rooms are taken into account and modified to set up the possible walking ways for an operator in the production line. The FME module *CenterLineReplacer* creates lines in the center of each defined polygon. The result of the *CenterLineReplacer* is a connected line through all corridors. If corridors are connected, also the lines are connected. In addition, the lines are separate if there is one separated corridor. After this creation of the center lines, the network must be prepared manually. This is done in ArcGIS. The lines are extended and connected to the equipment nodes. Also the main facilities for supply and delivery must be connected as well as the access nodes to the network and between the different floors or buildings. An extract of the created network can be seen in figure 5.6. The red line is representing the edges as possible walking ways of the network connecting the equipment and facility nodes as green points. The bigger red point is one access node which is a transfer node in the network.

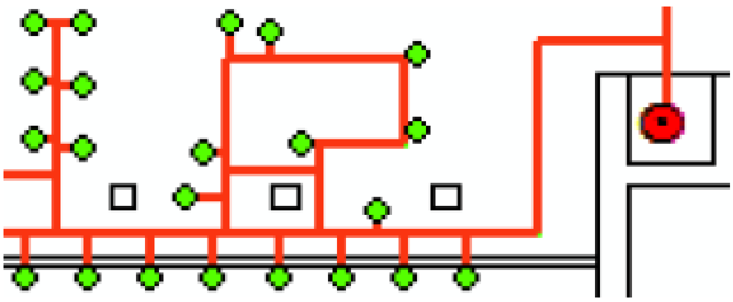


Figure 5.6: Extract of the created production line network.

As there are three Shapefiles with nodes that are necessary for one network, the three node files are combined to one Shapefile storing all nodes. This is done using the FME module *PointOnLineOverlay* which assigns each node of an equipment, facility or access node overlaying on the created edges of the network. The result are two Shapefiles one storing all relevant nodes and one all edges of the network. This is later on important to provide the opportunity to walk through the network. To enhance the routing function and to enable directions in the network, the assigning of a coordinate system is necessary. Therefore, a customized coordinate system was created. It is important to change the linear unit to millimeter, as the extracted lengths of the AutoCAD file are in millimeter.

To test the created network, a network dataset is created in ArcGIS with the Network Analyst. This enhances immediate a visualization in ArcMAP and a route can be created. The properties of the network dataset are that there are global turns allowed. That is important because an operator can turn on any place immediately. The connectivity policies for the network line belong to the end point and the connectivity policy of the network nodes to honor. Both files are in the same connectivity group and they are also the basis for the network. The elevation field is not used at the moment, because there is no continuous elevation field. The cost field of the lines is the length. The unit for the length is measured in meters as it is more readable than large millimeter values.
After the successful testing of the implemented network, the network nodes and network edges are inserted into the spatio-temporal database as separate tables using the Shapefile to PostGIS loader *shp2pgsql*. In the spatio-temporal database, the pgRouting extension is available and the edges and nodes have to be assigned to a network. Therefore, different views are created listing each node as either start or end node for a possible routing operation as well as a table to store the calculated route including the id of the start and the id of the end node, the id of the edge, all the geometries as well as the sequence. If the network is created in the database using pgRouting, it is possible to execute shortest path queries as for example the Dijkstra Algorithm to simulate the shortest path between equipments or facilities in the production line.

5.4 Creation of an Indoor Ontology for a Production Line

An indoor ontology of the production line environment is created using the free and open-source editor and framework Protégé. This framework provides an adaptable plug-in architecture to set up simple or even complex ontologies. The usage of rules is provided to enhance the software with the ability to solve a variety of different intelligent systems. Generally, an ontology is representing knowledge of a system in a hierarchical concept. Different types, relationships are possible within such an ontology, in this case the ontology of an indoor production line.
The created ontology for the indoor production line is including a navigation ontology on one hand and on the other hand the indoor production line environment and is relating to the work of Yang and Worboys (2011). In this case, the ontology describes entities of the indoor space which is defined as the production line environment as well as an integration of the navigation capability between the production line entities.
The characterization from sub-section 4.4.3 is taken into account. In the characterization the main entities in the production line environment are described as well as a depiction of the main entities that are necessary for the navigation or routing. As it can also be seen in section 5.3, a network based on a graph structure is created within the production line environment and will be used for the routing

purposes. Further information are concerning the production asset and the order of equipments. Hence, the equipments often perform a number of different processes in the production line and vice versa the order of equipments will be handled as production processes. The production asset has information about the possible transportation as for example the priority or how carefully the production asset has to be handled. Other information are concerning the quality and barriers affecting the transport of the production asset. A list of production processes is also provided to keep in mind the sequence of production processes.
Figure 5.7 shows an example of the developed ontology based on a barrier in the production line environment. A barrier is defined, to be either a constraint for the routing or it is the basic environment of the production line such as the wall. In the ongoing figure, the barrier is depicted to be fixed or moveable. A moveable barrier is defined as a production line facility or a contamination which occurs for a specific time period. The fixed barrier is covering equipments in the production line, the air quality which is limiting the transport or the wall of the manufacturing site.

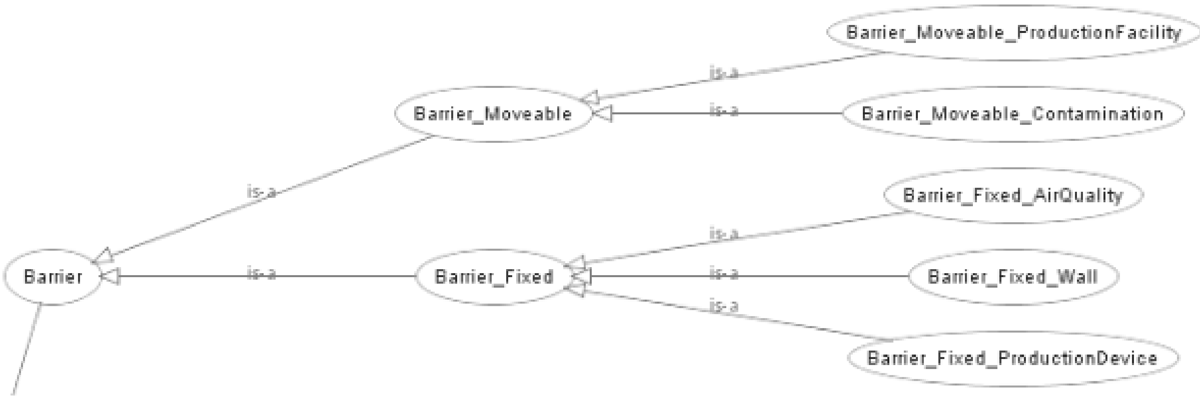


Figure 5.7: Barrier example of the developed production line ontology.

5.5 Extraction of Movement Paths

Next to the data preparation, data migration, creation of the network for the indoor environment is the extraction of the movement paths. This is possible, as the tracked position data is stored in the database and the equipments and facilities are linked to the network via a network node. During the development, four different types of movement paths are extracted. The movement paths are extracted using implemented PostGIS function and pgRouting in the background to calculate the shortest path between two objects according to the Dijkstra Algorithm according to one production asset. These four movement paths are:

- A direct connection from the used equipment to the next used equipment ordered by the time and the sequence of the recipe.
- A direct connection from the tracked positions for one production asset ordered by the provided timestamps.

- Connection between the used equipments along the created network by considering a timely order - *shortest path between the equipments*.
- Connection between the used equipments and the tracked positions between the equipments along the network by considering a timely order - *real path between the equipments*.

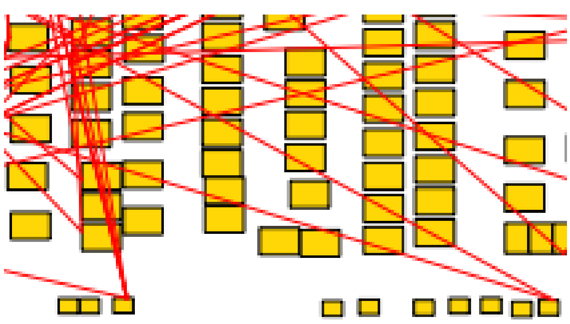
Direct Connection between the Equipment: This type of movement path is useful to give a general impression of the complexity in the production line. One production asset has to be selected, for which the path with the direct connection between the equipment has to be created. After the selection of a specific production asset, a PostGIS function is used for the extraction of the path. The PostGIS function is first of all selecting all the used equipment IDs for the production asset and orders them immediately regarding the check in timestamp. Next to the ordering, is the PostGIS function is calling a PostGIS function making a straight line between two points, which are at the beginning the first equipment and the second equipment and the line as well as the equipment information and timestamps are stored in an own table. This is repeated for the second and the third equipment and so on. Until the last direct line is appended to the table, it is possible to visualize the direct line connections between the equipments for example in Quantum-GIS for a further inspection An example of these tracks can be seen in sub-figure 5.8a.

Direct Connection between the tracked positions: An impression of the high number of tracked positions and the complexity of movements can easily be seen of one production asset. It can also be visually compared with the direct connections between the equipments of the same production asset. Another PostGIS function is selecting all the tracked positions for a specific production asset in a timely order. Next to it, with a provided function in PostGIS the direct lines between the tracked positions are created and stored in an own table for further visualization purposes. Sub-figure 5.8b shows such a visualization in Quantum-GIS

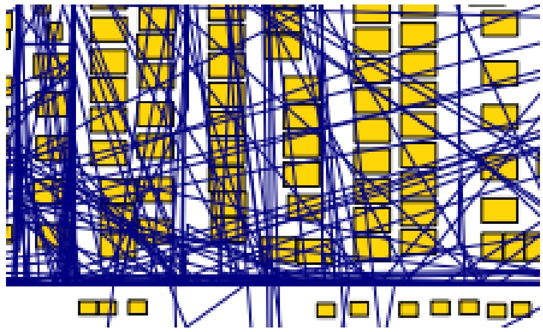
Shortest Path - Equipments along network: The basis for the creation of the shortest path between each equipment in the timely order along the created network is also a function selecting and ordering the used equipment for a specific production asset. As the equipment has an equipment node which is linked with the network it is possible to close the gap between the equipment and the network. This linkage between equipment, equipment node and that each equipment node is a network node is described in the previous section 5.2 and section 5.3. A function is developed, determining the network node ID of each used equipment and is appending the timestamps from the check in and the check out of the production asset. Next, for each movement or sub-route from one equipment to the next in the timely order the shortest path algorithm of pgRouting along the network is executed. The result of each sub-route is stored in one separate table which stores the overall route from the first until the last equipment. Additionally, the table stores information about every start and end node, the timestamps check in and check out, the used edges of the network and their sequence as well as the length of each edge and the an identifier of the production asset. All this is useful for further analysis and calculations. It is possible to calculate the length of the shortest path between all production steps, to create a time slider visualization according to the check in and check out or even to identify the most used edges or hot spots of used edges. The shortest path and the used equipment can also easily be visualized by connecting Quantum-GIS to the PostGIS database and load the specific tables. Sub-figure 5.8c shows a path along the network in the production line with standardized and anonymized equipments and an extract of the table storing the described information.

Real Path - Equipments and tracked positions in between: Similarly to the extraction of the shortest path for a production asset, the real path is extracted using a PostGIS function selecting all relevant equipments for a specific production asset ordered by the time. The next step is a function starting to proof if there are any tracked positions before the first equipment. If there are tracked positions, the shortest path algorithm is executed from tracked position to tracked position until the first equipment occurs regarding to the time. After the check out at one equipment, the check in at the next equipment is identified and all tracked positions in between these to timestamps are identified. The shortest path algorithm is combining the equipment with all the tracked positions till the production asset is processed by the next equipment. This is repeated until all equipments are connected along the network including the tracked positions before or after each equipment. An important issue is the implementation of a nearest neighborhood function, which is necessary to map the tracked positions to the network. The tracked positions are situated in the middle of the corridors and walking paths and therefore, a nearest neighborhood function between each tracked position and the network nodes is implemented. For the whole extraction process of the real path can be for example that the routing process starts at equipment node. After the check out of the equipment the check in at the next equipment is identified and the id of the node is stored. All tracked positions in between are identified and each tracked position uses the nearest neighbor of the network nodes through the implemented nearest neighborhood function. The route goes from the first equipment node to the tracked positions until the next equipment appears and so on. All the sub-routes from equipment to tracked position, tracked position to tracked position, tracked position to equipment or equipment to equipment are stored in a separate table. The table includes the used edges and the sequence of the edges, the timestamp of the tracked position or the check in and check out at a device, all the IDs of the network nodes and the geometry enabling a further analysis as it is described for the shortest path. Sub-figure 5.8c shows such a route along a network and a table storing the described information.
Figure 5.8 shows different extracted movement paths. In sub-figure 5.8a the direct connection between equipments can be seen as red lines. In contrary, sub-figure 5.8b connects the tracked positions with blue lines. The last sub-figure 5.8c shows a path along a network which can either be a shortest path or the real path as a red

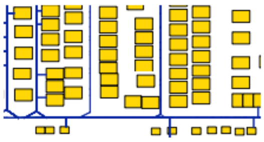
line through the production line. The attributes are also highlighted which are the length of the edge, the start and end id of the routing task and the sequence of the edges for each task. The start time and end time for the routing as well as a count of how often the specific edge is used within the overall path. The last attribute shows which type of connection it is as for example equipment to equipment.



(a) Direct connection between used equipments.



(b) Direct connection between tracked positions.



edge_pk	length	start_id	end_id	seq	starttime	endtime	count	type
189	1.3591000564	768	778	0	2013-06-20 13:36:14	2013-06-20 13:36:15	1	Device to LotPosition
187	3.1735	778	818	1	2013-06-20 13:36:14	2013-06-20 13:36:15	1	Device to LotPosition
170	3.17540000119	819	818	2	2013-06-20 13:36:14	2013-06-20 13:36:15	3	Device to LotPosition
171	3.5375	820	819	3	2013-06-20 13:36:14	2013-06-20 13:36:15	5	Device to LotPosition

(c) Extracted path along the network and the attributes in the table.

Figure 5.8: Extracted movement paths in the production line.

5.6 Data Mining and Data Analysis

The data mining and data analysis section describes the training process of the SOM using the described SOMatic training library in section 3.2.3. After the training process SOMatic viewer is used to enable the comparison of the created component planes of the SOM. Therefore, also small applications in the programming language Processing are created including the SOMatic library to simplify the visualization and comparison of the various component planes. The last sub-section of the data mining and data analysis section considers the physical and attribute space of the input data. Therefore, a website is developed including a Pr ocessing JavaScript application. This application enables the visualization of the SOM, the different component planes projected on the SOM, the analysis of Best Matching Units projected onto the SOM as well as the time related visualization supported by a time-slider.
Generally, the data mining is executed on two different datasets. The first dataset is taking into account the used equipment for each production asset. The second dataset covers the quality issues occurring during the overall production and the used equipment affecting these quality issues. The analyzed datasets described in detail are:

Production Asset and the Equipment: A dataset is created according to the second data model in sub-section 4.4.4. This approach gives the opportunity to point out the frequency of each equipment per production asset. In this case, the sequence itself is not really considered, but the similarity of the overall sequence is comparable by the frequency of the equipment. Therefore, the described model builds a matrix of the production asset and the equipment - Matrix{production asset x equipment}. This means for the SOM a dimensionality of more than 400 dimensions, as the dimensionality is considered to be the attributes or the columns which will be used for the training processes. Data for more than a few thousand production assets is available and has to be brought into the modeled form for the training of the SOM. For each production asset the appearance of the equipment has to be determined to show the absolute frequency of the equipment for one production asset. To enable a comparison between the frequency, a normalization is necessary for each equipment, SOM dimension or column of the data. These column normalization is done by a MIN-Max normalization assigning the highest value of the column the number 1 and the lowest value is assigned to 0. This normalization is a crucial issue to ensure a comparison between the component planes of this SOM and the Best Matching Units of the production asset projected onto the SOM.
The dataset is prepared using PostgreSQL/PostGIS database and psql-Functions. A temporary table is created in the specific form that each equipment is a column and one column is storing the identifier for the production asset. Psql-Functions are used to count the appearance for each equipment for the production asset and a 0 value is assigned if the equipment was never used in the sequence of one production asset. After that, the dataset is stored as a Comma Separated Value file and is column normalized in Microsoft Excel. This was done separately, to ensure and to proof the normalization of the high number of dimensions of the SOM and the number of production assets. Equipment which is never used for any production asset is deleted from the dataset and not considered for the training of the SOM, as the component planes would be empty and each neuron would have the same value.

Quality Issue and the equipment: The input dataset for the quality issue and the equipment is created according to the definition of the first approach described in sub-section 4.4.5, where a quality issue is defined as a sub-sequence of equipment. This approach is selected, as one of the main advantages is that each equipment can have a certain concern triggering a quality issue. An example is already described in sub-section 4.4.5, as an equipment can have a huge impact on the quality of a production asset but another equipment is triggering the quality issue with only a little impact on the quality but it reaches a defined threshold of a measurement value. Also these equipments are considered in this approach, as they appear in several sub-sequences and gives the equipment a specific importance that has to be considered. This leads to the possibility to identify equipments that are more likely to be involved in a quality issue. Further, the selected approach has some kind

of a dynamic length of equipment per sub-sequence referring to the occurring of the quality issues if a production asset has only a few or many quality issues. This varying length helps to point out equipments that have a certain impact on quality issues. Finally, this general approach can be applied to other production lines easily and is not fixed or limited to one specific type of production line environment and production assets.
Similarly to the dataset of the production asset and the equipment the second data model of sub-section 4.4.4 is used showing the frequency of each equipment. In this case, each quality issue is creating a subset of the overall sequence. Each subset leads to one quality issue and through the frequency of the occurring equipment assumptions are possible. Therefore, the described model builds a matrix of the quality issue and the equipment - Matrix{quality issue x equipment} defining a high-dimensional space for the analysis. This means again a SOM with a dimensionality of about 400 dimensions and a higher number of quality issues than production assets which are a few thousand.
This dataset is also created using the PostgreSQL/PostGIS database with two temporary tables. One table is simply delivering basic information about the quality issue as for example which was the last equipment, at which time the quality issue occurred and a unique identifier. The second temporary table uses the described format of a column for each equipment and one column for the unique identifier of the quality issues. A psql-function is implemented to extract the sub-sequences for each quality issue and stores the values in a separate table, to enable a comparison of the sub-sequences and proof if the right equipment is selected. After that, another psql-function is used to count the appearance of the equipment for each quality issue. Finally, the data is exported as a CSV file and each column is normalized between a 0 and 1 range similarly to the dataset of the production asset and the equipment. Equipment which is never used within any sub-sequence of a quality issue is deleted to decrease the number of dimensions as there is also no opportunity to gain new knowledge.
Additionally, the third approach can be applied by considering the statistical values how often an equipment triggers a quality issue. This involvement of a quality issue can be normalized by the number of production assets that passed through the equipment and how often the equipment triggers a quality issue. By this method a percentage value can be calculated and enables a possible comparison of equipment. This will also be done by the help of the PostgreSQL/PostGIS database, where it is also possible to extract production assets and quality issues at a period of time as for example one day, one week or a month. After this it is possible to compare the spatial distribution of the percentage values and identify areas with less quality issues and areas with a higher amount of quality issues considering the time. Therefore, database views can be created and such a view can be visualized in for example Quantum-GIS very easily.

5.6.1 Training Process of the SOM using SOMatic

To train the SOM, the SOMatic Training Software is used which is described in subsection 3.2.3 for each of the prepared datasets. The general training process starts with the input data which is described before as either the dataset of the frequency of the equipment per production asset or the frequency of equipment per quality issue. Each dataset has to be trained separately and the result is a *.dat file storing the input vectors for the SOM and the codebook *.cod file storing the SOM and the component planes.

After starting the Java application of SOMatic using an integrated development environment like eclipse for Java it was possible to load the input dataset in the form of a *.csv file. Next to the loading of the base data the global settings have to be defined such as a normalization method, the normalization range for the values or what happens with missing values. No normalization method was selected as the input data is already normalized before due to computational performance that occurred during the normalization with the SOMatic Training Application. The normalized data is loaded and the attributes important for the training of the SOM have to be selected. This means each attribute has to be selected except the unique identifier, which is important as a reference back to the real data and is not relevant for the training of the SOM. Next to the successful loading and processing of the data is the actual training process.

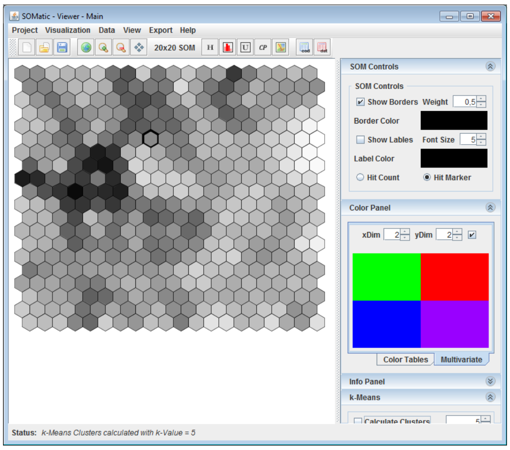
To train the SOM, it is necessary to prepare an initial starting SOM. Therefore, the topology of the SOM has to be selected as well as the initialization method. In this case, a hexagonal topology is selected which provides more accurate information showing the similarities and clusters as there are more neighboring neurons than using a rectangular topology. A randomized initialization method is selected for the SOM at the very beginning of the training process. The size of the SOM is defined as to be a 20*20 SOM. That means, the number of neurons is in X-direction and Y-direction 20 neurons, resulting in overall 400 neurons.

The training process is divided into two separate phases, whereas the resulting codebook file of the first training phase is the input an input file for the second training phase. According to Skupin and Esperbé (2011), the two phase training method is recommended, as the first training method determines a broad and global structure and the second phase goes into detail into regional and local structures. Further, the two phase training decreases errors occurring during the training process of a SOM. The training parameters for the first phase and the second phase are a neighborhood function of bubble and a Euclidean measure of the similarity between neurons and training vectors. The difference between phase 1 and phase 2 is the number of training runs, a different learning rate and the neighborhood radius.

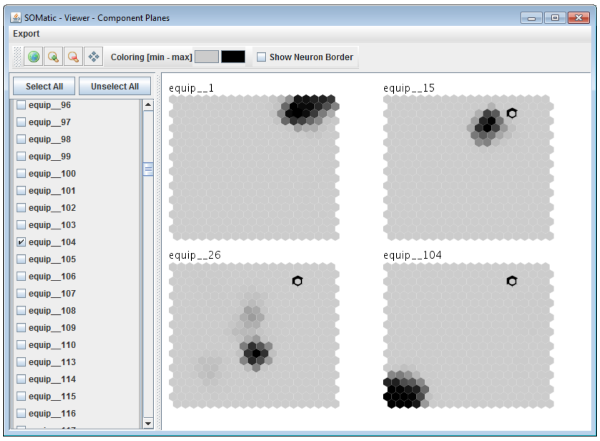
The first training phase focuses on the global structure of the SOM. Therefore, 25.000 training runs are set and a learning rate of 0.05. 25.000 training runs means, that each input vector is used at least two times. The neighborhood radius is defined as to be 20, so each neuron of the whole 20*20 SOM is considered. In comparison, the second phase starts with the codebook file of the first phase and has 100.000 training runs with a learning rate of 0.04 and a neighborhood radius of 5. After the training of the second phase the SOM can be visualized and the component planes can be compared and analyzed. The parameters are the same for both of the input datasets.

5.6.2 Comparison of Component Planes using SOMatic

After the creation of to SOMs it is possible to analyze or compare the different component planes. Basically, this is done using the SOMatic Viewer application. For a dynamic comparison and creation of clusters dynamically via a slider, a Processing sketch is implemented including the SOMatic library and the SOMatic functionalities. The SOMatic Viewer application has to be adapted to the new input data as there is a static coding behind this functionality. After the adaption, it is possible to load the before created input vectors and the codebook file into the SOMatic Viewer Application. Similarly to the SOMatic Training application, the Viewer can be started in the integrated development environment eclipse. After the start up, the selected 20*20 SOM is loaded and displayed in a pre-defined color scheme. Abilities are provided to switch between different visualization methods such as a *u*-Matrix visualization, the Component Planes, a mapping of the Best Matching Units or a *k*-Means cluster calculation of the SOM. A color schema can be set for the component planes or the standard visualization from white for low values to black for high values can be selected. The loading of the created SOM is computational expensive by considering a few thousand production assets or quality issues and the high number of dimensions, as SOMatic is calculating the BMUs on the fly as well the calculation of the clusters is done for each step. The SOMatic Viewer application is also used to create the component plane catalogs in Appendix B for the production asset and the equipment and Appendix C for the quality issue and the equipment. Figure 5.9 shows the frame of SOMatic with the opportunity to compare the planes. It can easily be seen, that it is possible to select or unselect component planes via buttons or simply the check boxes. In addition, specific component planes concerning the equipment can be selected manually by the user. Another feature is the hit score of the main frame, where specific neurons can be selected and the BMUs are listed in a small info panel. This feature enables a simple comparison of the neuron in for example the *u*-Matrix showing the Euclidean distance between the vectors of the codebook file. The main frame with the u-Matrix visualization and a selected neuron can be seen in sub-figure 5.9a. In addition, four component planes are selected in sub-figure 5.9b, showing the same neuron highlighted.



(a) Main Window of the SOMatic Viewer Application.



(b) Comparison of Component Planes using the SOMatic Viewer Application.

Figure 5.9: SOMatic Viewer Application.

5.6.3 Analysis of Physical, Attribute Space and Time based on Processing JavaScript

To analyze the created SOMs a website is developed including the Processing JavaScript visualization language. Processing JavaScript is one part of the Processing visual programming language and enables the interactivity of visual elements with standard elements on a Website such as for example jQuery elements. A website is created, as there is the opportunity of an easy distribution or access in a company and in case of the interaction of standard web-elements with the SOM. This interaction is an importation issue and will be used for the exploring of the created SOMs and the layout of the production line. The equipment defines the physical space, the SOMs the attribute space and the time is considered through a time slider.

Until now, Processing JavaScript is not able to execute any external Java libraries as for example the SOMatic library. This leads to the issue, that the CORE-library of Processing has to be used. Therefore, a standard processing sketch is created, which is the first prototype. The basis for the prototype is provided by Dr. Andrè Skupin and is loading basic SOM including the input vector file and the codebook file. This sketch is displaying the SOM and is extended with the ability to project the BMUs onto the SOM. After this visualization, the whole Processing Java sketch is adapted and transformed to Processing JavaScript. For example, an own Table Class was developed to store the information due to a lack of data types and functionalities. Another example was, that Processing JavaScript was not able to handle a BufferedReader, which has to be replaced. In the end, the final first prototype was a simple Processing sketch in JavaScript loaded from a website displaying one SOM and the physical space was displayed as a scalable vector graphic.

The second prototype was enhancing the first prototype, which was the website, with functionalities to interact with the SOM. Therefore, a time-slider and a drop-down list is implemented using the jQuery javascript library. Additional information is necessary to synchronize the time-slider with the projecting of the BMUs. This prototype is developed with the SOM about the production asset and the equipment. At this point, already all component planes are loaded into the program and one component plane is displayed. The component planes are linked to the drop-down list, where the user has now the ability to switch between the component planes of this SOM. The time-slider will be able to handle the projecting of the BMUs onto the SOM. Each production asset has a start and an end point of the production process. If the selected time on the slider falls into this range, the BMU of the specific production asset is projected onto the SOM. Hence, more input information is necessary to handle the interaction correctly. In figure 5.10 it can be seen, that timely information concerning the production asset has to be loaded as well as the codebook file and the input vectors to visualize the SOM. Due to computational performance on the website, the BMUs are calculated in the Processing Java sketch and stored into a file and loaded in the Processing JavaScript sketch. The website is loading four different files to visualize one SOM in a proper way including the time-slider and drop-down list. The input files for the SOM are tab-stop delimited.

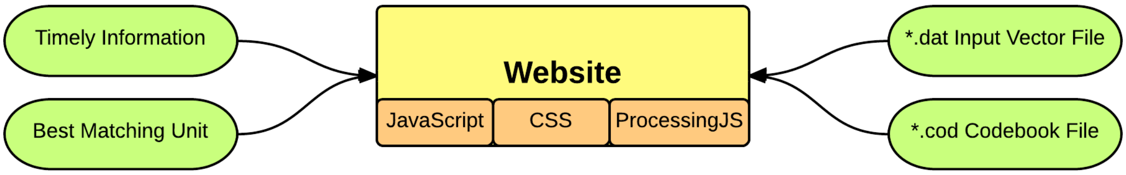


Figure 5.10: Different input files for the visualizing the SOM on the website and projecting the BMUs accordingly.

The website is finally implement with both SOMs next to each other. The quality issues as well as the production assets are only projected onto the corresponding SOM according to timestamp and time range of either the quality issue or the production asset. The physical space is according to the equipment of the production line and the affected equipment by a quality issue will be highlighted. All these is implemented in Processing JavaScript, JavaScript and HTML. Summarizing, prototypes are developed to enable the way from Processing Java to Processing JavaScript. Thus, it is possible to interact with the before created SOMs as the projecting of the BMUs is linked to a time-slider and component planes can be changed by an additional control element, a drop-down list. All this enables an analysis of visual patterns on the SOM that can occur either in the physical space, in the attribute space on the SOM, in both spaces or even according to the time.

Chapter 6

Results

The result section includes the developed indoor ontology of the production as well as the implemented network. The network enhances the production line with the routing ability to analyze the movement behavior and to compare different possible tracks. Another described result a comparison of possible tracks in the production line. Finally, the spatio-temporal data mining is executed and the results are on one hand a possible comparison of component planes of the equipment throughout the production line of the tracked production assets or the quality issues, as they are defined in sub-section 4.4.5 of the modeling phase. Component Planes are analyzed and compared of the frequency of equipment per production asset as well as the frequency of equipment and the quality issue. Last, a website is covering the analysis of the physical space, attribute space and time.

6.1 Spatio-Temporal Database

The spatio-temporal database is a database model based on timestamping to store the spatio-temporal data. Generally, the spatio-temporal database consists of two connected parts. These two parts are on one hand the general information of the production line concerning the production asset, equipment, facilities and further tasks and on the other hand the connection to the network which has to be included within the database. Figure 6.1 shows the relational database concept which is applied in PostgreSQL/PostGIS. The general spatio-temporal data concept and attributive information is stored in the tables with the green header and the tables with the blue header are the basis and connection for the network, which will be stored in this schema. Further, the tables highlighted in figure 6.1 are described under the figure.

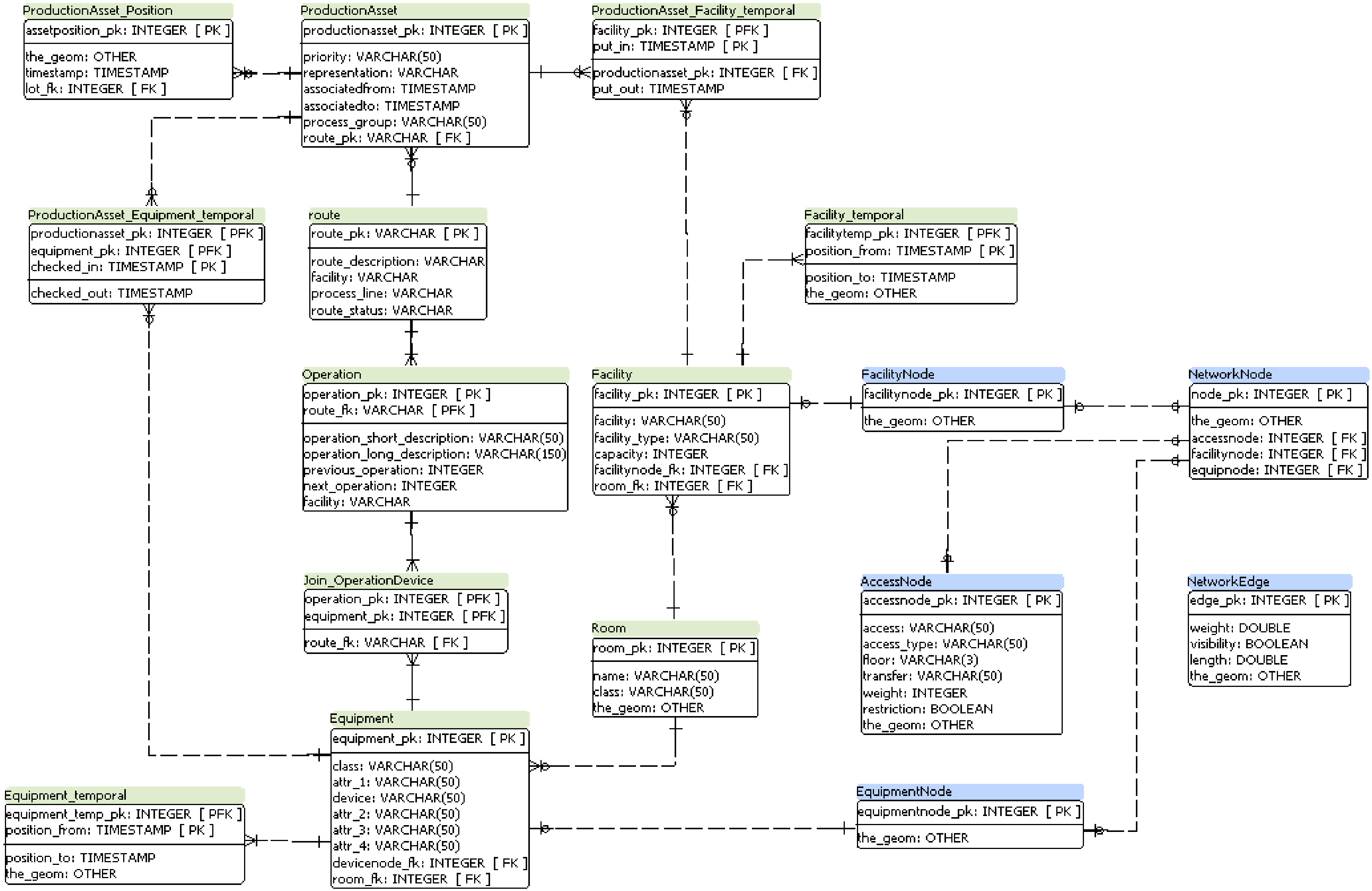


Figure 6.1: Spatio-temporal data concept including the network.

ProductionAsset is the main table in the spatio-temporal database. The production asset is the transported object in the production line environment. Therefore, attributive information is stored such as the priority, timely associations concerning the overall time, representation tasks such as names as well as a possible classification of the production asset. The production asset is linked to a changing position, a route describing the sequence of equipment, the connection of the production asset with the equipment and a facility.

ProductionAsset_Position is storing the changing position of the production asset throughout the production line. This is possible by storing the relationship to the production asset, the geometry and the provided timestamp of the position.

Equipment are the entities processing the production assets. Therefore, they are separated into different classes and types of equipments, the name and descriptive attributes. This table is also linked to a temporal component, the connection of the production asset with the equipment, relationship to the network, a room and possible operations of the equipment.

Equipment_temporal covers the position and geometry of the equipment. This is based on two different timestamps storing the time when the equipment was build up at a position until the position is changed. Hence, historical queries and questions can be answered.

Facility is basically a shelf or a table in the production line. This includes a type to distinguish between the type of facility and a name for the identification. A capacity is added where it is possible to add information how many production assets can be deployed on or in a facility. The facility is linked to a room, the network via a facility node, the temporal component of the facility and finally the connection if a production asset is put into a facility.

Facility_temporal is very similar to the equipment temporal as it is storing the changing or moving position of the facilities over time. This is possible by the help of two timestamps, the geometry and identifier to link the attributive information of the facility with the spatio-temporal information.

ProductionAsset_Equipment_temporal this table is necessary to be aware of the timely aspects occurring if a production asset is processed by an equipment. In this case, it is important to know information about the active equipment, the processed production asset and the timestamp when the production asset checks in and checks out at the equipment.

ProductionAsset_Facility_temporal the same task as the table before ProductionAsset_Equipment_temporal is established with this table. Here information can be stored if a production asset is put into a facility or is taken out. Therefore, two timestamps have to be provided and the identifier of the production asset and the facility.

Room is used to divide the production line into smaller areas and rooms. The room has an own id and helps mainly to identify equipment and facilities standing in the same room or not.

Route is including attributive information about the sequence of equipment to finalize a production asset successfully. Routes are also divided into different groups or classes and include several operations. Therefore, the route is linked with the table operation and the main production asset table.

Operation describes the processes that can be executed at an equipment and are necessary for the production asset. Each production asset has a specific sequence of operations due to the predefined route.

Join_OperationEquipment is connecting the operation and the equipment. Basically, it is simple many-to-many connection as an equipment can execute different types of operations.

FacilityNode is necessary to link the facility to the network and enable the routing functionality. Therefore, during the data preparation a node is created for each facility.

EquipmentNode is storing a node for each equipment in the production line. The node is created in the data preparation and migration phase.

AccessNode is storing another node information occurring in the production line. Access nodes include the connectivity between indoor and outdoor geography, indoor geography and indoor geography, elevators or stairs. Therefore, restrictions have to be provided as well as a weight which can be the time dependency of an elevator.

NetworkNode is an important table which is later on necessary for the creation of the routable network using pgRouting. It is summing up the three types of nodes which are facility node, equipment node and access node.

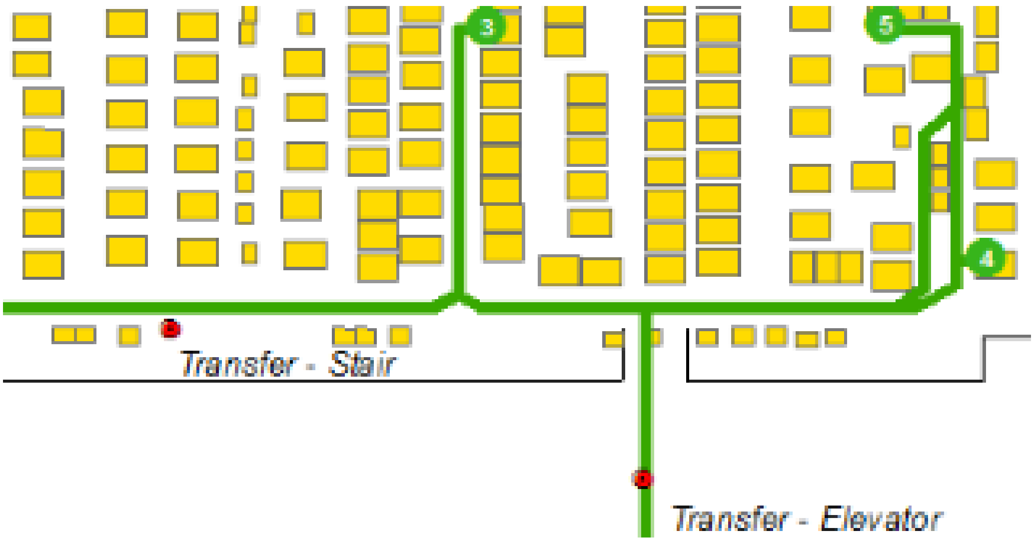
NetworkEdge is a table including the edges of the network. Therefore, the geometry is an important issue, a restriction if it is possible to walk or to drive a transport cart, a weight as well as the visibility of the edge in the network.

6.2 Routable Network in the Indoor Environment

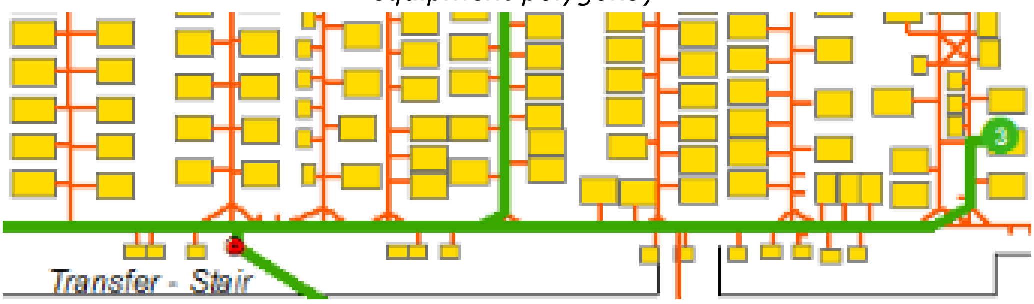
A created network enables the production line with a routing functionality to establish the opportunity to compare possible paths in the indoor environment. In section 5.3 the creation is described using ArcGIS and the FME Workbench. The routing capability is evaluated in ArcGIS with the help of the Network Analyst. Therefore, a specific sequence of equipment is manually selected and the shortest path is displayed. Figure 6.2 shows two different examples of routes in the production line. One is considering the constraint that a stair is not allowed and once it is allowed to use a stair.

Sub-figure 6.2a shows a snippet of a route through the production line. In yellow standardized equipment polygons are displayed and the proposed route of the

ArcGIS network analyst is a green line with a numbering of the different selected end points. Red points are highlighting the defined access nodes, which are in this case a transfer over the stair and a transfer using the elevator. The route itself starts on the left side along a corridor going straight ahead until a right turn until the selected point 3 is reached. After that, the route goes the same way back until the corridor is reached again and the route turns left and goes more less straight until end point 4. From end point 4 the way goes up to the equipment selected as end point 5. After the last selected point in this area the route goes back onto the corridor and turns left to the elevator, which has to be selected as stairs are not allowed in this example.
The second example in sub-figure 6.2b shows as well the standardized polygons in yellow, another proposed route in green and the access nodes. In contrary, the developed network is also displayed as red lines in the background to show the variety of walking ways and the stair can be used for the calculation of the shortest path. This route also starts from the left side and visits selected equipment end points in their order as end point 3 is for example in the right corner of the sub-figure. After that, the way goes back along the corridor until there is a left turn and the route picks the stair for the transfer between different floors as it is allowed in this example route.



(a) Example route through the production line using the elevator (*standardized equipment polygons*).



(b) Example route through the production line using the stair (*standardized equipment polygons*).

Figure 6.2: Different example routes through the production line.

6.3 Indoor Ontology of a Production Line

The developed indoor ontology is closely related to the work of Yang and Worboys (2011) and extended with entities highlighting the indoor space of a production line environment and an integrated navigation ontology based on a graph. The ontology is depicted in figure 6.3 and the different entities and concepts are described within this section. Further, the developed indoor ontology covered in this result section is also a forthcoming paper in the proceedings of GIScience 2014 in Vienna (Scholz and Schabus, 2014).

Production Unit: A production unit represents the whole equipment of a production line. For example a Facility a Device or a Equipment that are used during the various production steps. The subclasses are *ProductionUnit_Facility* and *ProductionUnit_Device*.

ProductionUnit_Device: A device is the production unit used for the processing of goods. The device has a fixed position in the production line.

ProductionUnit_Facility: The facility supports transport processes in the production line. The goods can be placed on shelves or tables if they are waiting to be processed or transported. The subclasses of a facility are ProductionUnit_Facility_Moveable and ProductionUnit_Facility_Fixed.

ProductionUnit_Facility_Moveable: A moveable facility is used to support a high stock of goods in the production line. They are e.g. bottleneck shelves used to store an extra amount of production assets. Such objects are removed if the stock in the production line is decreasing.

ProductionUnit_Facility_Fixed: Fixed facilities represent tables, shelves and further not moveable equipment in the production line.

Barrier: A barrier is limiting the transportation or movement behavior in the production line. The subclasses are *Barrier_Fixed* and *Barrier_Moveable*.

Barrier_Fixed: A fixed barrier is limiting the movement behavior and cannot be changed very easily. Subclasses are *Barrier_Fixed_Wall*, *Barrier_Fixed_ProductionDevice* and *Barrier_Fixed_AirQuality*.

Barrier_Fixed_Wall: A wall is a fixed barrier. It is limiting the transport behavior within a production line.

Barrier_Fixed_ProductionDevice: The device in a production unit is linked with several infrastructure items such as electricity and gas lines and is regarded as a fixed or not easily changeable barrier.

Barrier_Fixed_AirQuality: For several production goods the air quality in a clean room is of importance and is also a barrier for the transport and movement behavior.

Barrier_Moveable: Moveable barriers represent mainly barriers that can change over time very easily. The subclasses are *Barrier_Moveable_ProductionFacility* and *Barrier_Moveable_Contamination*.

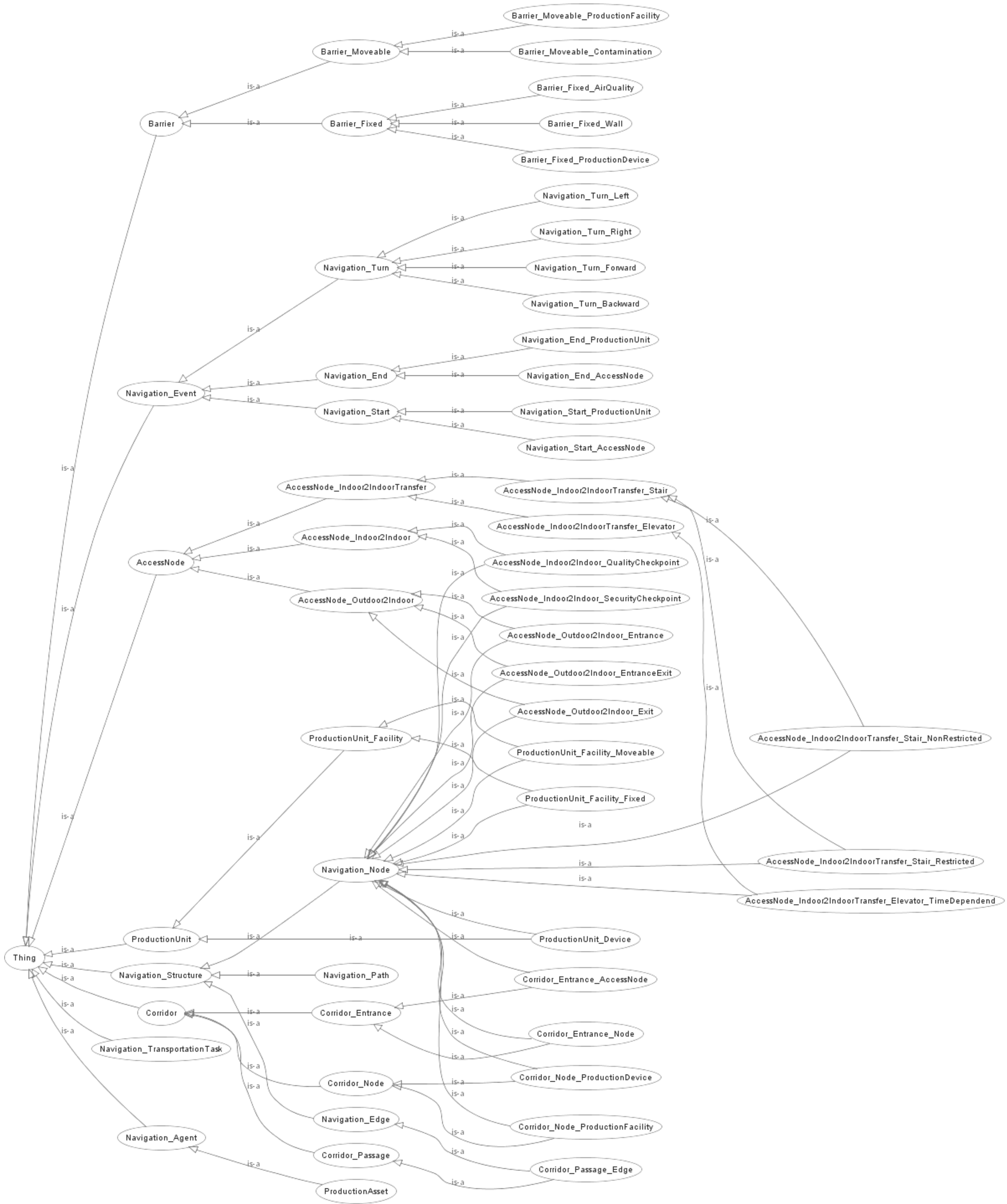


Figure 6.3: Indoor Ontology of the Production Line Environment including the Navigation.

Barrier_Moveable_Contamination: A contamination is a barrier over time. Hence, a certain production good is not allowed to enter a specific area of the production line.

Barrier_Moveable_ProductionFacility: Any production facility can impede movement as it is limiting the space for transportation. E.g. The position of shelves may easily be changed if they are not necessary any-more.

AccessNode: An AccessNode is linking outdoor and indoor space or vice versa. The subclasses are *AccessNode_Outdoor2Indoor*, *AccessNode_Indoor2Indoor* and *AccessNode_Indoor2IndoorTransfer*.

AccessNode_Outdoor2Indoor: The connection from outdoor geography into the indoor environment. Therefore, the subclasses *Entrance, Exit and Entrance-Exit* are necessary.

AccessNode_Outdoor2Indoor_Exit: The exit is representing the way from an indoor geography back to the outdoor geography. This is necessary as there exist designated doors for leaving a production line (especially true for production environment with clean rooms)

AccessNode_Outdoor2Indoor_EntranceExit: The EntranceExit represents both the way from outdoor geography to indoor geography and backwards.

AccessNode_Outdoor2Indoor_Entrance: The entrance enables the interaction and movement from outdoor into the indoor space.

AccessNode_Indoor2IndoorTransfer: The transfer indoor is representing the connection in the same indoor space, thus connecting e.g. different floors.

AccessNode_Indoor2IndoorTransfer_Elevator: The transfer of production assets with an elevator in order to change the floor level.

AccessNode_Indoor2IndoorTransfer_Elevator_TimeDependend: The time dependence of an elevator is used in order to integrate the average waiting time until an elevator is available, due to the fact that elevators are mostly not available instantaneously.

AccessNode_Indoor2IndoorTransfer_Stair: A stair enables the transfer between different floors in an indoor space.

AccessNode_Indoor2IndoorTransfer_Stair_NonRestricted: Traversing a stair is allowed for all production asset types.

AccessNode_Indoor2IndoorTransfer_Stair_Restricted: The traversal of a stair is not allowed for certain production asset types.

AccessNode_Indoor2Indoor: This class represents the transfer between different indoor spaces – e.g. different production halls.

AccessNode_Indoor2Indoor_QualityCheckpoint: A quality check such as an e.g. air quality check with an airlock.

AccessNode_Indoor2Indoor_SecurityCheckpoint: The entrance to certain areas can be restricted.

Corridor: A corridor is describing and including the ways where an operator – i.e. human being – can walk and transport the production goods in the production line. The subclasses are *Corridor_Node, Corridor_Passage* and *Corridor_Entrance*.

Corridor_Node: Corridor nodes include the starting point, end point or interaction point of a navigation process.

Corridor_Node_ProductionFacility: A start point, end point or interaction point can be a production facility. For example a good has to be brought to a shelf because something has to be controlled.

Corridor_Node_ProductionDevice: A production device is mainly a start or end point for the transportation or navigation as the production goods are processed here.

Corridor_Passage: The passage itself is representing the way in between the navigation task.

Corridor_Passage_Edge: An edge is used between the different nodes and is combined to a passage along the corridor.

Corridor_Entrance: Corridors need entrance points to the network for navigation and transportation in the production line.

Corridor_Entrance_AccessNode: The access node is one opportunity where operators or production assets are accessing the transportation network.

Corridor_Entrance_Node: Entrance nodes can also be Production devices or facilities.

Navigation_Event: Any navigation task is described through the classes *Navigation_End, Navigation_Start* and *Navigation_Turn.*

Navigation_End This class represents the destination of a transportation or navigation task.

Navigation_End_AccessNode: An access node is the destination node of the navigation process if e.g. a production asset leaves the production line.

Navigation_End_ProductionUnit: The transportation between devices or facilities implies that a production facility or device is the end of the navigation task.

Navigation_Start: The navigation start is representing the start of a navigation task, which can either be an *AccessNode* or a *ProductionUnit*.

Navigation_Start_AccessNode: An access node is the start of the navigation if a production asset is entering the production line.

Navigation_Start_ProductionUnit: The production unit is a starting point for the navigation.

Navigation_Turn: During the navigation a production asset can perform several actions. These actions are the subclasses *Navigation_Turn_Right, Navigation_Turn_Left, Navigation_Turn_Backward* and *Navigation_Turn_Forward.*

Navigation_Turn_Right: The production asset turns right.

Navigation_Turn_Left: Represents a turn to the left.

Navigation_Turn_Backward: This event is a turn backward or represents backwards moving.

Navigation_Turn_Forward: This is a move forward.

Navigation_Agent: The agent that is navigating through the indoor space. **Production_Asset:** This class represents the navigation agent, and encompasses various types of production assets with different properties that have an influence on the suitability of a certain route and the choice of a certain route. **Navigation_Structure:** This class contains generic entities that are necessary for route calculation proposes. A sequence of instances of the subclasses Navigation_Node and Navigation_Edge on which an agent moves defines a Navigation_Path. The objects of the class Navigation_Structure are help to specify the indoor space entities in terms of representation in a graph with nodes and edges.

6.4 Comparison of possible Tracks in the Production Line

The comparison of different possible tracks in the production line is possible through the extracting of movement paths described in section 5.5. According to this section, there are four different opportunities whereas the first two opportunities are a direct connection between either equipment or tracked positions. The second two opportunities are the so called real path considering the equipment and tracked positions in between along the network and the shortest path considering the equipment along the network. By inspecting the direct connections between the tracked positions it is already possible to ask specific questions about what happened to the production asset while it was moving from one room to the next room and it took a long time, but the rooms are next to each other. Another result is the overall comparison of the real path and the shortest path for the same production asset by considering the total length or the frequency of specific edges in the network. It is also possible to compare the real path and the shortest path for the same production asset from equipment to the next equipment in the sequence and to investigate what happened, is the real path already optimal or is there a potential to decrease the transportation way. Summarizing, there are opportunities to compare the real path and the shortest path by considering the overall distance, by comparing the way from equipment to equipment of the real path and the shortest path and also the difference of the frequency how often an edge was traversed.
First thoughts can be given by looking onto the direct connection of each tracked position ordered by the provided timestamp. This shows basically, the complex movement behavior of the production asset throughout the whole production steps.

In figure 6.4, a small extract of the production line can be seen where the equipment is colored in yellow and the direct connections between the tracked positions are straight blue lines. The data is derived from the spatio-temporal database and visualized using Quantum-GIS. As it can be seen, the selected red line in the figure highlights the change from one room into the next room which takes along amount of time for such a short movement. The information window in Quantum-GIS allows an exploration of the timestamps. Therefore, questions can be stated such as why was the production asset waiting for the next equipment; why does the transport took so long; what happened to the production asset during the time; were other production assets also waiting? This shows a variety of questions that can be stated by a simple visual inspection.

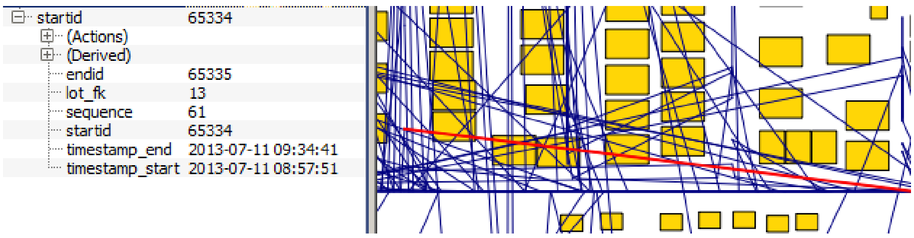
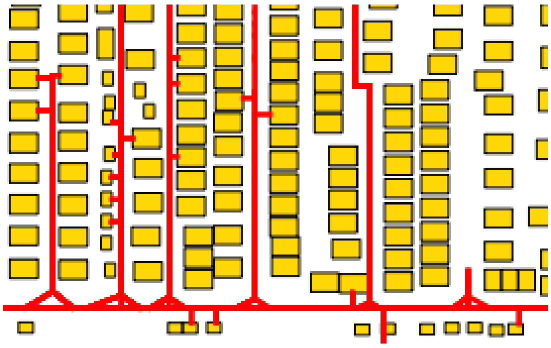
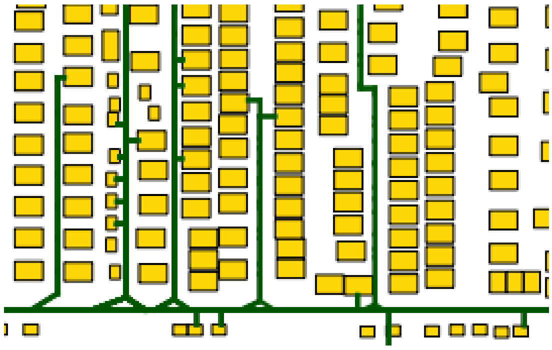


Figure 6.4: Direct combination of tracked positions for one production asset.

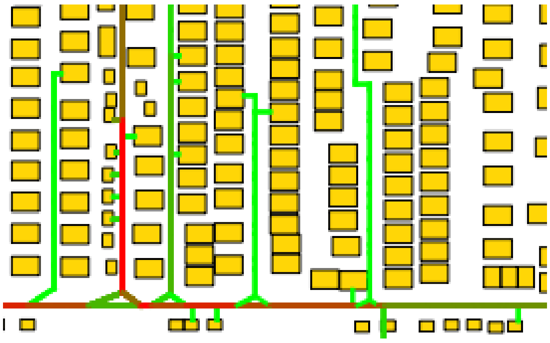
Second is the comparison of the described real path and the shortest path including the frequency how often an edge has to be traversed. Therefore, figure 6.5 shows both ways with the same visualization parameters and an example how the frequency is highlighted. The equipment is again illustrated using standardized yellow polygons. The real path is visualized using a red color and the shortest path a green color. For the frequency visualization a color schema from green to red is used, which describes the number how often a specific edge is used. Green edges mean the edge was not used frequently and a red edge highlights often frequented edges within the network. Often frequented edges can help to identify potential bottle necks within the production line. Sub-figure 6.5a shows the real path through the production line of one randomly selected production asset with an overall distance of a few kilometers. In addition, sub-figure 6.5b shows the shortest path through the production line for the same randomly selected production asset. In the real path, the maximum frequency of an edge is nine times the maximum frequency of shortest path. This can for example be an issue, if the production asset changes the position in one room very often, because it has to wait for the equipment and the operators are carrying it around. An assumption can also be made about the difference of the real path and the shortest path, were the shortest path is 3 times less than the real path. A statistical comparison of the tracks is carried out but not included in this thesis, as it is taking into account security and privacy issues of the manufacturing site providing the data. Sub-figure 6.5c shows the frequency of traversing an edge. The green color represents a low frequency of one production asset at the edge and red a high frequency.



(a) Real path through the production line.



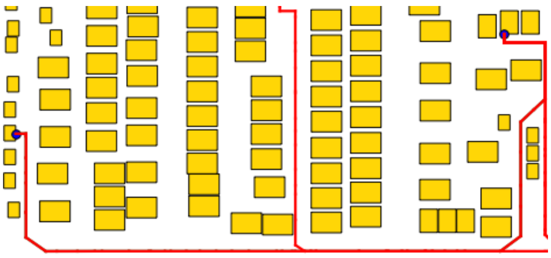
(b) Shortest path through the production line.



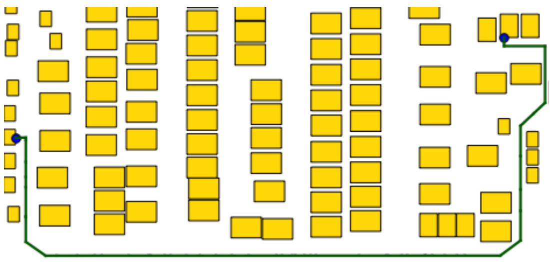
(c) More and less frequented edges.

Figure 6.5: Different example routes through the production line.

The last comparison of possible tracks is a direct comparison of the real path and the shortest path between the way from equipment to equipment. An example of the difference can be seen in figure 6.6a where the same visualization method is used as before there are yellow standardized polygons for the equipment and the trajectories are visualized as either green lines for the shortest path or red lines for the real path. Sub-figure 6.6a shows the real extracted path from equipment to equipment with a distance of approximately a bit less than 200 meters. In sub-figure 6.6b the shortest path between the same two equipment is a bit more than the half of the distance. This can also be seen by comparing the actual paths in the figure as the transport in the real path goes to other rooms and other equipment first until the target for the selected production asset is reached.



(a) Real path from equipment to the next equipment.



(b) Shortest path from equipment to the next equipment.

Figure 6.6: Difference of the paths between one equipment to the next equipment.

6.5 Component Planes of the Production Asset and the Equipment

The component planes of the production asset and the frequency of the equipment offers the opportunity to compare different equipments regarding their similarity or dissimilarity concerning the production assets. To be able to compare the component planes, it is important to understand the model standing behind the input data for the training process of the SOM. The different data models are described in sub-section 4.4.4 and section 5.6 describes the selected data model and the argumentation for the data model is pointed out. To show how it is possible to interpret this type of component planes some examples are described, which should lead to enough knowledge to interpret the various component planes listed in component plane catalog in Appendix B.

The first example in figure 6.7 shows two different component planes of equipment 28 and equipment 137 in the production line. Both component planes represent a 20*20 SOM and the coloring represents dark or black with high values and low values with a gray color. As it can be seen, the compared component planes are mostly different, with one area in the left top corner. In this area, equipment 28 has very high values and also equipment 137 has middle values occurring in that area. All other areas are very distinct, as for example the very dark area in the middle of the component plane of equipment 137 is some sort of cut out in the component plane of equipment 28 and shows that there is a big dissimilarity in this specific area. Basically, production assets are processed by the equipment in the production line and on the component planes of this model, it is possible to see which equipments occur most likely together in the sequence of a production asset or not. By comparing the two equipments with further knowledge of the spatio-temporal database it can also be seen, that equipment 28 is used by about 500 different production assets and most of the time it is used only once. Equipment 137 is used by more than 3500 different production assets and it is often used several times as the number of carried out processes is more than 5000. These values lead to the assumption, that equipment 28 is used less frequently than equipment 137. This can also be seen in the component planes as the component plane of equipment 28 is mainly focused in one area, whereas the other component plane is distributed over the whole SOM and used frequently regarding to the darker values. Finally, both equipments are also used together by about 300 production assets and the attributive data gives also insight, that equipment 28 is used mostly one and equipment 137 is used more often if both equipments are used together. Summarizing, equipment 137 is used more frequently by different production assets as the values are distributed over the whole component plane and equipment 28 is focused on specific production assets. As there is an overlapping area, sometimes they are used in the same sequence.

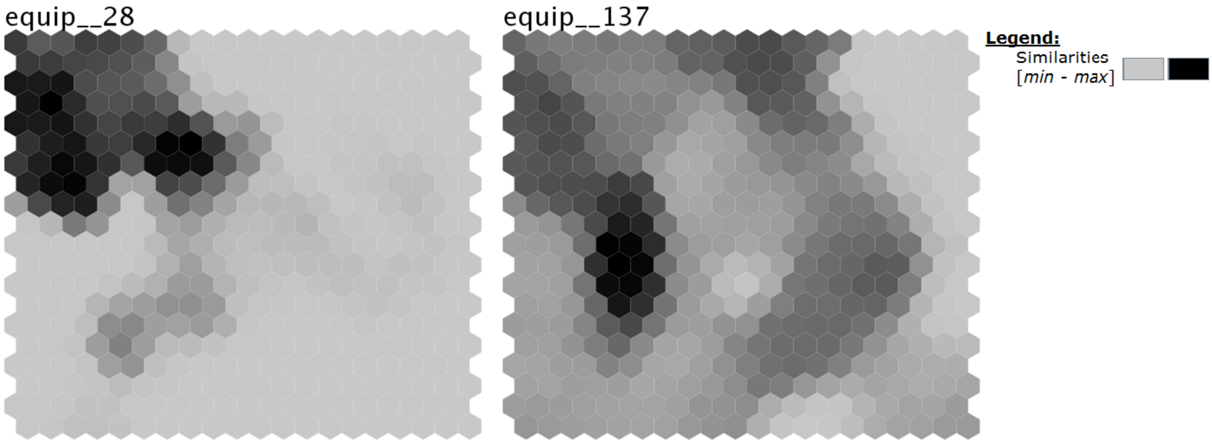


Figure 6.7: Comparison of two component planes (Equipment 28 and 137).

Another possible comparison of component planes can be seen in figure 6.8 including equipment 21 and equipment 392. At the first point of view it can already be clearly said, that equipment 21 and 392 are not used together, as their neurons with dark values are not similar and further apart. Additionally, the component plane of equipment 21 has an interesting pattern resulting in several smaller and distinct areas over the SOM. Two areas are visible with darker values showing a higher frequency of production assets in this area, than in the other areas with lower values. All in all, equipment 21 has to be compared with other component planes. If another equipment has dark and high values in the same area, it can be said that these equipments are often used in the same frequency. The two dark areas lead to an assumption that this equipment is frequently used in two different sequences of production assets and sometimes it appears in other sequences not as frequently reflecting the middle gray values.

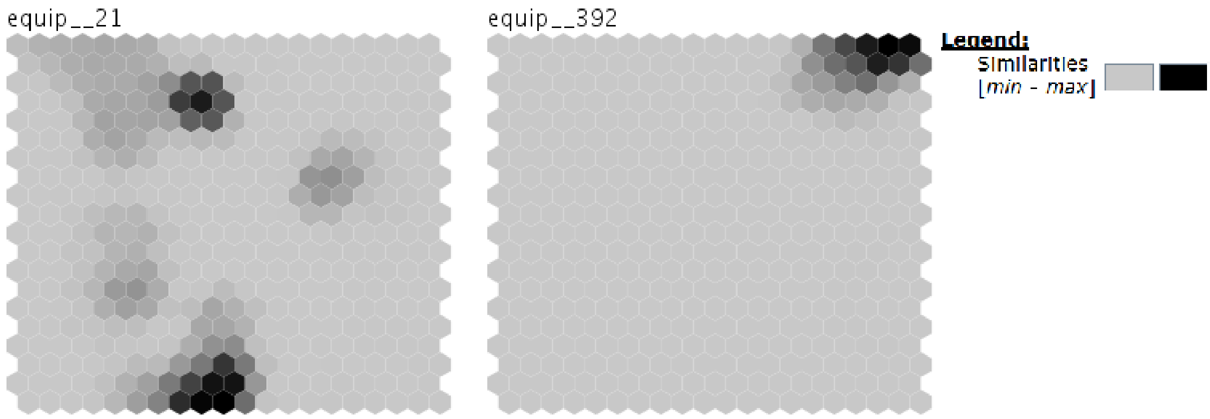


Figure 6.8: Comparison of two component planes (Equipment 21 and 392).

6.6 Component Planes of the Quality Issue and the Equipment

The component planes of quality issue and the frequency of the equipment offers the opportunity to compare different equipments regarding their similarity of the quality issue. To be able to compare the component planes, it is important to understand the model standing behind the input data for the training process of the SOM. The different data models are described in sub-section 4.4.4 and section 5.6 describes the selected data model and the argumentation for the data model is pointed out. To show how it is possible to interpret this type of component planes some examples are described, which should lead to enough knowledge to interpret the various component planes listed in component plane catalog in Appendix C.
The first example off the quality issue and the used equipment in figure 6.9 shows to similar equipments. As it can easily be seen, similar neurons occur in both of the selected component planes. Further, equipment 28 is more frequently used if this quality issue is occurring as more darker areas can be identified. By comparing the values in the spatio-temporal database, equipment 28 is used in about 500 sequences of quality issues and equipment 34 in about 600 sequences. Both equipments occur most of the time only once an the sequence of a quality issue. A production asset has two quality issues and both quality issues are using both equipments on a different day with the same equipment triggering the quality issue. This can lead to an assumption, that it is very likely if the production asset uses the equipments again, another quality issue can occur. By simply comparing the neurons of both component planes, it defines that the equipments are looking very similar concerning the quality issues. A main finding, is that in neurons where both equipments have very dark values, it is likely that the same quality issue can occur. The distribution of the clusters in both neurons show, that the equipments are used in the sequence of different quality issues and some are occurring more likely than other quality issues. Therefore, the developed website provides the ability to project the quality issues onto the SOM.

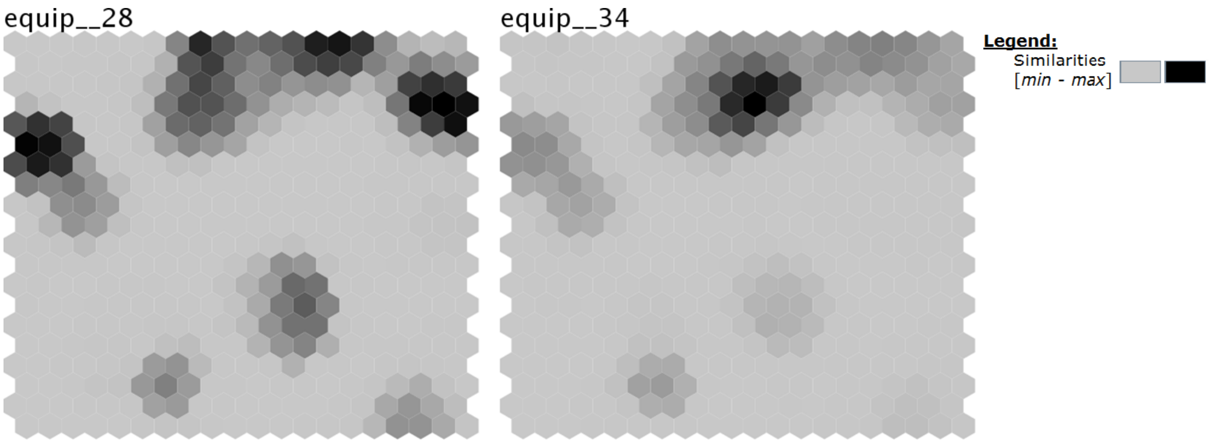


Figure 6.9: Component planes of the quality issue(Equipment 28 and 34).

The second example shows quality issues concerning equipment 2 and equipment 367. These two component planes are completely different regarding the clusters of affected neurons. By comparing the component plane with information provided in the spatio-temporal database, it can also be seen that the equipment are in two different categories as it is supposed to be. Further, equipment 2 has a very limited cluster of affected neurons in the left-bottom corner. Due to the limitation of neurons, the equipment is only appearing rarely in the sequence of a quality issue. In addition, equipment 367 appears in several different sequences of quality issues as there are some distinct and distributed areas. Further, equipment 367 appears in one larger cluster of quality issue sequences in the right-top corner more frequently than in the other clusters.

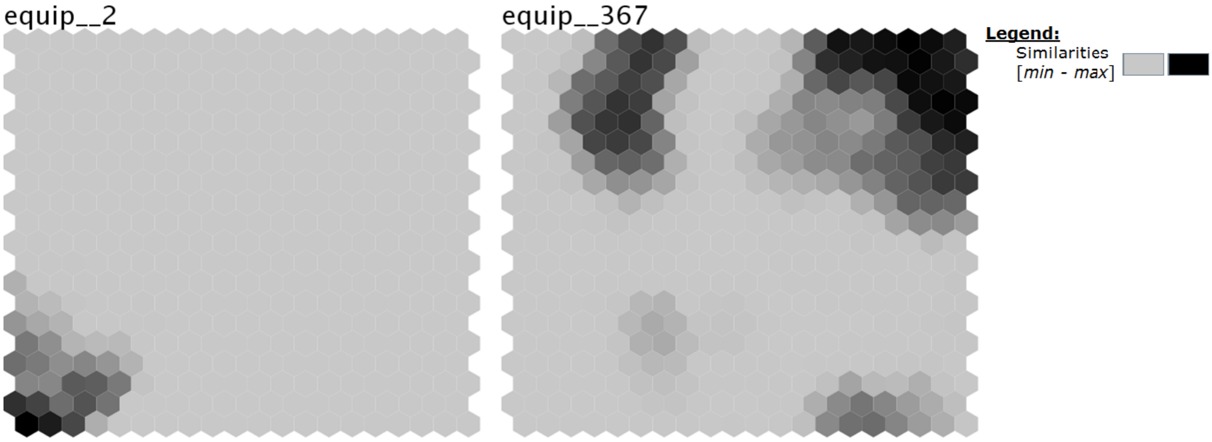


Figure 6.10: Component planes of the quality issue(Equipment 2 and 367).

6.7 Interactive Website combining Self-Organizing Maps, Physical Space and Time

Finally, a website is created including the created SOMs and the additional physical space including the production line layout. Through drop-down lists it is possible to change the component planes projected onto each SOM. Another important issue is the timely component, which is associated by a time-slider in the range of the selected input data of the production assets and the quality issues. Due to security issues concerning the data, only numerical identifiers are projected onto the SOMs representing either the production asset or the quality issues, which can be linked to their name or further information through this identifier. Another security issue is the mapping of the physical space or the production line. The production line is mapped through a various number of standardized polygons. According to that, the website is also not shown completely only extracts and examples of the physical space are provided within this result section.

In figure 6.11, the left SOM is displaying the SOM regarding the production assets and the equipment and on the right side the quality issue and the equipment. As it can be seen in both drop-down lists, the component plane for equipment 137 is displayed. The component plane is illustrated by drawing red circles into the center of each affected neuron. This means in detail, if a neuron has a high value the size of the circle is getting bigger and if there is a small circle there is a low value for the selected component plane. If a neuron is empty, it has simply no value for this component plane. At the start up, the time-slider is at the starting position and no BMUs are projected onto the SOMs. By hovering over a neuron, the dominating dimension is highlighted. This means, the equipment with the highest value in the neuron is appearing as long as the mouse is over a neuron.

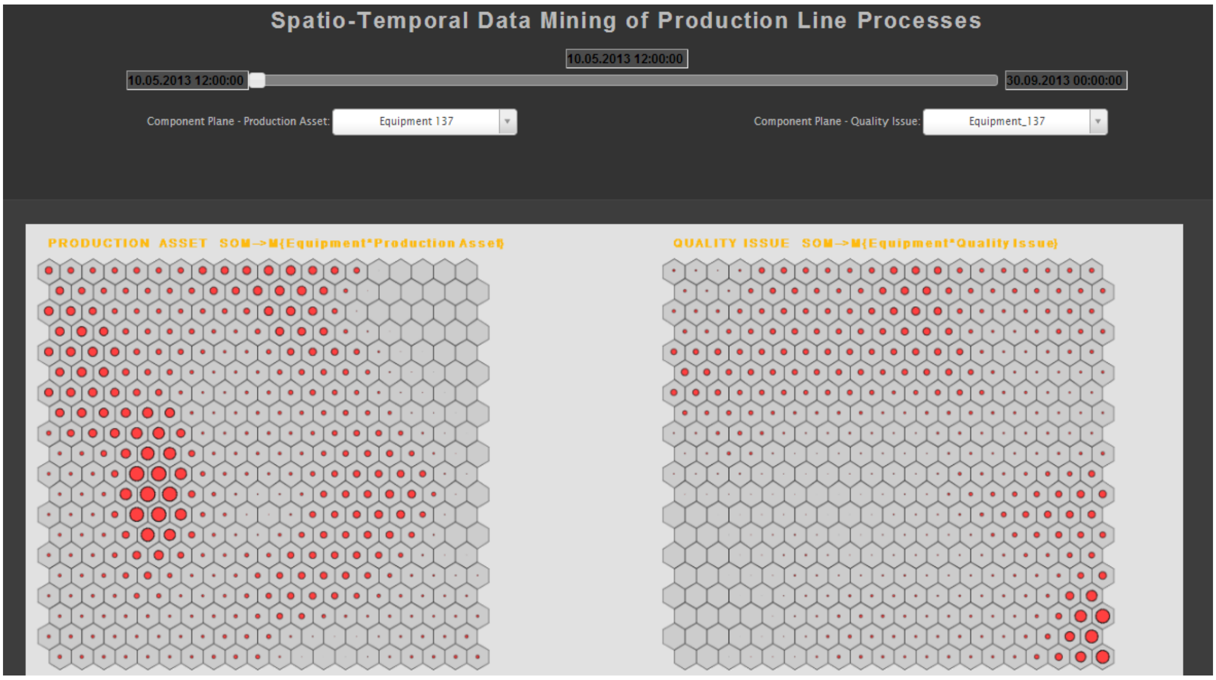
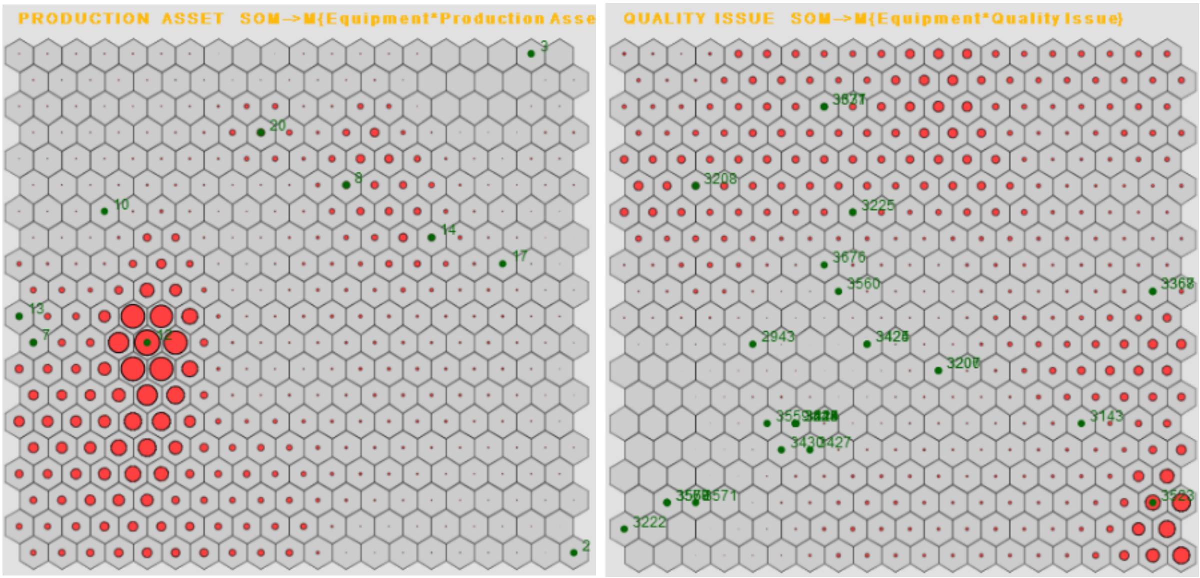


Figure 6.11: Website including the control elements and the two SOMs.

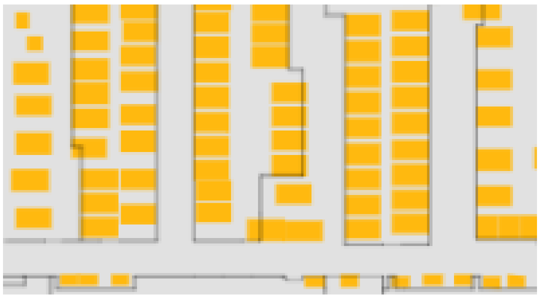
Additionally, figure 6.12 shows the two SOMs in detail with projected BMUs onto the SOM. Sub-figure 6.12a is including the SOM of the production asset and the equipment. For this SOM, the component plane for equipment 176 is displayed. It can easily be seen, that this component plane has a cluster of high values around the BMU of the production asset with the identifier 12. Further, the values are spread over more or less the whole same and points out that this equipment is used very frequently in sequences of many production assets. For the production asset with the identifier 12 is it used very often at is in the cluster with the highest appearance of this equipment. Sub-figure 6.12b shows the quality issues projected onto the corresponding SOM. The component plane in the background is the component plane for equipment 137 and shows high values in the right-bottom corner with one quality issue projected into this area at a particular timestamp. The quality issue with the id 3523 is very likely to use the equipment 137 within the sequence until a quality issue occurs. A cluster of quality issues can be detected in the left lower side of the SOM, where four neighboring neurons have projected BMUs into it. It also has to be considered, that each quality issue is projected into the middle of the neuron and if this neuron is the BMU for more quality issues they are somehow covering each other. In this case, it defines that there are at least four very similar BMUs close to each other. After identifying such a cluster it is possible to investigate these quality issues with the help of the spatio-temporal database. This can lead to new knowledge about similarities of quality issues and the affected equipment. Further, it is possible to take a look into the physical space and proof at which equipment the quality issues occurred. As the SOM shows a certain similarity in the attribute space it is also possible that there is a similarity in the location.



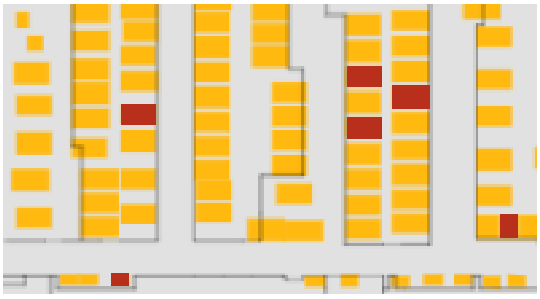
(a) Projected Production Assets on the SOM **(b)** Projected Quality Issues on the SOM.

Figure 6.12: Projecting BMUs onto the SOM to a certain timestamp.

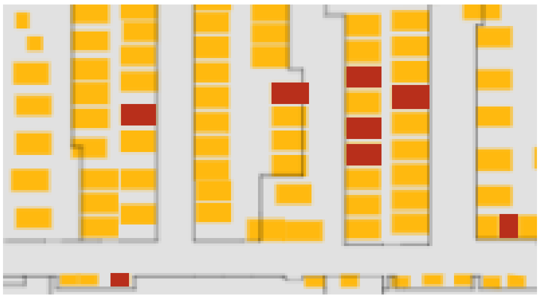
The physical space is limited to the equipment and displays an equipment in a dark red color, if this equipment is affected by a quality issue at the timestamp selected with the time-slider. Therefore, an example pattern is created to show the evolving of such a pattern over time. A simulation is done, due to the limitation of the selected extract of the production line as well as to the standardized and randomized polygons. If a quality issue occurs, a time interval of one hour is defined in which time period the BMU and the equipment is highlighted to gain the opportunity to identify any patterns. In figure 6.13, a demonstration of such a pattern in the physical space can be seen at four different timestamps. Sub-figure 6.13a shows the selected extract of the production line with the standardized polygons in yellow at the start up of the website. By using the time-slider BMUs are projected onto the SOM on the website and if a quality issue occurs the triggering equipment is highlighted in a red color which can be seen in sub-figure 6.13b. At this timestamp, no clear pattern can be identified as the affected equipments look randomly distributed in the production line. Next sub-figure 6.13c shows already an area with many affected equipment in a very close distance or even next to each other. Directly, it is possible to see in which area the quality issues are occurring and if they are also similar in the attribute space over the quality issue SOM. The last sub-figure 6.13d shows already that this identified pattern is not existing anymore regarding to the time constraint. This gives an example, how it is possible to investigate the physical space, the attribute space of the two SOMs by the help of a time-slider. If a pattern is detected, the spatio-temporal database in the background provides a user with further attributes that can be considered and gives inside into the pattern.



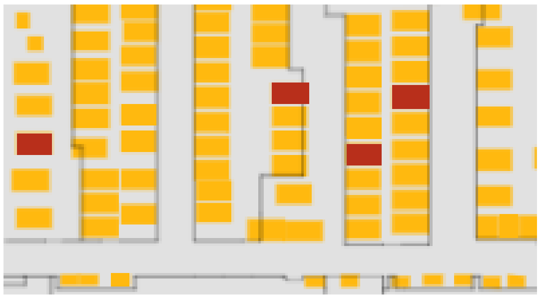
(a) Timestamp 1.



(b) Timestamp 2.



(c) Timestamp 3.



(d) Timestamp 4.

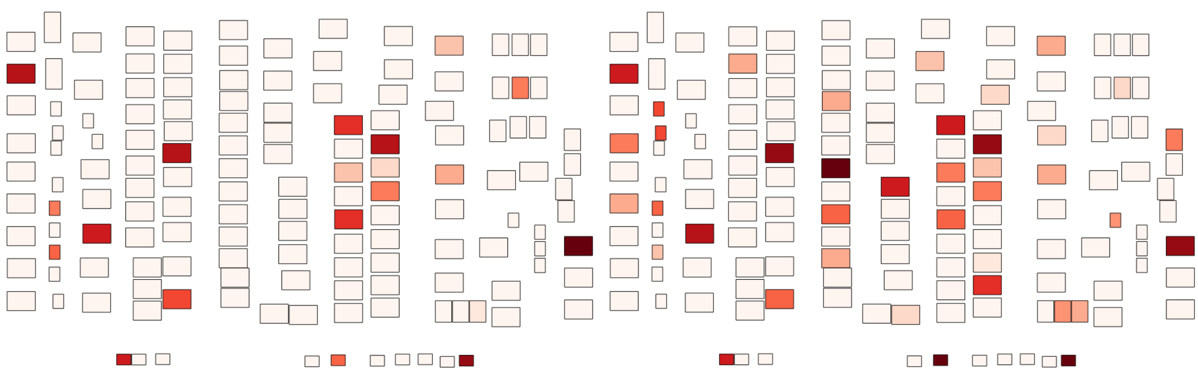
Figure 6.13: An evolving pattern of the equipment in the physical space.

To sum up, a website enables a user to investigate the attribute space which is including one SOM about the production asset and the frequency of the equipment and one SOM about the quality issue and the occurring equipment. All this is implemented in a website and the interaction is enabled through a time-slider and drop-down lists to change the component planes projected onto SOM. The time-slider is handling the projecting of the BMUs onto the SOMs regarding their time. The production assets are mapped according to their overall existence and the quality issue depend on a time range which is 30 minutes before and 30 minutes after their recording. In the physical space, the equipment triggering quality issues are also highlighted regarding this time range of the quality issues. All this helps to investigate patterns over space in the physical space, similarities in the attributes and time using the time-slider.

6.8 Comparison of Quality Issues in the Physical Space

An additional possibility of a quality issue is also defined in sub-section 4.4.5, which is the third described approach where a quality issue is defined by the equipment triggering the quality issue. To remember, this quality issue is useful if for example a order of specific objects is proofed at each equipment. The implementation is briefly described in section 5.6 as the values are prepared in the PostgreSQL/PostGIS database and visualized using Quantum-GIS.

This approach, enables a direct comparison of equipment in the production line by computing the percentage value of quality issues per equipment considering each operation for an equipment and the occurring quality issues. Therefore, a time range can be defined enabling an analysis of the physical space over for example a few hours, a day, a week or even a month. Figure 6.14 shows such a created map of the physical space visualized in Quantum-GIS using standardized polygons for the equipment. As it can be seen, the equipment is visualized using a red color schema with a changing saturation to differentiate between low values with a low saturation and higher percentage values with a darker saturation. Sub-figure 6.14a shows a visualization of the percentage of quality issues occurred during the overall number of processes in one day for the equipment. This approach applied on a whole day can lead already to knowledge about spatial patterns such as problem with the communication between the RFID receiver in an area or a contamination in an area can be identified or monitored. By applying this quality issue on the data of a whole week in sub-figure 6.14b it is possible to identify for example evolving patterns of equipment in an area by comparing ongoing weeks. In this case it is possible to identify equipment without a spatial pattern but with high values, which shows that there it would be possible to proof the equipment as many processes fail. Another example can be a long term pattern, where a whole area combining several equipments next to each other are effected with a medium to high number of quality issues. By monitoring several weeks or days in a consecutive time, patterns can be identified such as rooms or areas affected for example by air quality or contamination problems.



(a) Quality issue per equipment in the time range of one day.

(b) Quality issue per equipment in the time range of one week.

Figure 6.14: Quality Issues per Equipment in a Time Range.

Chapter 7

Discussion

Within this section the results of the research are discussed. At the beginning, the developed spatio-temporal data structure and database are reviewed. Further, the indoor production line is analyzed and leads to the created graph based network which is reviewed, as it is the basis of the comparison of movement paths. Then, thoughts about the modeling of the base data for SOMs are made, as they are essential for the final results and the interpretation. Finally, general thoughts about the analysis of the component planes and the developed website are given.

Spatio-Temporal Data Structure and Database

The developed spatio-temporal data structure and spatio-temporal database are the basis of this research. Therefore, different spatio-temporal data models are considered in the literature review such as the snapshot model or an event-oriented data model. As the availability of the input data is proofed including the implemented indoor positioning solution and the production line environment, the data model based on time-stamping is select. It fits to the provided input data as all spatio-temporal aspects are provided as either one single timestamp or two timestamps defined as a start and an end time. In the end, this data model successfully stores and provides the data in a useful way. Hence, the data preparation and migration is a semi-automatic task that has to be executed taking a long time considering the various input features.

Graph Based Network and Indoor Navigation Ontology

Based on the provided input data and the spatio-temporal database a graph based network is created including an indoor navigation ontology for a production line. Such an ontology looks different than current indoor navigation approaches as for example described by Yang and Worboys (2011). The difference is that there are other entities than standard rooms or doors as the production line under review is for example in a cleanroom environment. The navigation ontology describes the indoor space as well as the affordance based navigation task of a production asset. A routing methodology is applied in a prototypical application. Thereby, a graph based network enables the routing functionality and creation of movement paths in the production line environment. By proofing this concept of mapping positions to a graph based network, 41097 tracked positions are evaluated and 97.3% are lying within a 1 meter buffer of the created network. This shows , that the created network fits to the production line environment. Movement paths of the production assets are extracted including tracked positions and the equipment. This leads to an assumption of the path as it is described by Jeung et al. (2011).

SOMs for the Analysis of the Movement Behavior

As SOMs are created for the analysis of movement behavior and quality issues, several possible data models are described to show possible similarities of the sequences of production assets. The selected data model for the training process of a SOM is including the count of each equipment for each production asset. This describes the frequency of equipment showing if a equipment was used often by a production asset or never. As a normalization method for the equipment is applied it is possible to compare the frequencies of different production assets showing if they are similar or dissimilar. Regarding this similarity, assumptions can be made concerning the similarity of the sequence as no sequence is stored and used for the training. To consider the sequence, a different data model has to be used creating either much more dimensions for a SOM which will be computational expensive or far more features that have to be visualized. In this case, other visualization techniques have also to be considered, which will lead to an overload of data that has to be visualized. An example can be a similarity matrix consisting of more than 10.000 features, setting up a Matrix of 10.000 * 10.000 rows and columns and 100.000.000 items making it very hard to identify single patterns.
When comparing the component planes of the SOM, it is possible to identify similar equipment for the production asset or quality issues. Examples of how to compare the component planes are stated in the result section. Further, a component plane catalog is creating in the appendix for both types of component planes. In addition to the component planes, it is possible to add data to gain a deeper inside and more knowledge about similarities and specific equipment.

Website to compare the Physical Space, Attribute Space and Time

In addition, the developed website is a first prototype enabling the visualization of a SOM on a website with the usage of Processing JavaScript. This prototype implementation has the opportunity to interact with the SOM by standardized JavaScript features such as a drop-down box to handle and project component planes. As the time-slider is moved and the relating BMUs of the production asset or the quality issue are projected onto the SOM, problems can occur due to the high number of production assets. The high number of production assets and the long timely range of the production asset, the BMUs look very crowded onto the SOM and it is hard to identify patterns. There is a higher number of quality issues than production assets but with a timely range of at least one hour and not a whole week it can be seen that it is an appropriate opportunity to explore the data. The website is also linked with the physical space of the production line and the equipment of a quality issue is highlighted to identify spatial patterns. Due to security issues of the company providing the data and the production line layout, standardized and modified data is used. This starts by standardized and randomized polygons in the physical space and ends by mapping of identifiers and no names on the website.

Nevertheless, this approach shows the ability of spatio-temporal data mining of movement behavior and quality issues by the use of SOM. This includes the setting up of a spatio-temporal data structure, an indoor environment, the creation of a network and the extraction of movement paths. All this leads to a website enabling an exploration of the created and prepared data.

Chapter 8

Summary

This section summarizes essential findings and results of this research. Additionally, basic ideas for future work and improvement of the research is presented.

8.1 Conclusion

This research represents the development and implementation of an indoor production line environment with a routing functionality building the basis for spatio-temporal data mining of the movement behavior and quality issues in a production line. The main results, are a spatio-temporal data structure and database storing all information used in the further analysis, the development of an indoor navigation ontology based on graph structure and the comparison and extraction of different movement paths. The final result, are SOMs including either the production asset or the quality issue linked to the physical space represented by the equipment including a timely component visualized on a website enabling a user interaction.

Spatio-Temporal Database

The first major result is the spatio-temporal database developed for the input data. The developed spatio-temporal database considers a database model based on time-stamping and is implemented using a PostgreSQL/PostGIS database. This spatio-temporal database is a central point during the whole analysis as it is storing all the data including the spatio-temporal data, attributive data and also the network which enhances the routing functionality. Therefore, the pgRouting extension of PostGIS is used to calculate for example shortest paths with the help of developed psql-Functions. Furthermore, an ontology is created describing the whole indoor environment including the entities of a production line as well as the navigation itself. Through this ontology, a graph based network is created and stored in the spatio-temporal database. In addition, different movement paths are extracted using psql-Functions and are also stored in the spatio-temporal database to be able to analyze the movement behavior. To sum up, the first main finding is a spatio-temporal database putting together all the provided input information, the whole spatial components of the production line and the developed network through the production line.

Analysis of the Movement Behavior and the Quality Issue

The second main finding is regarding the analysis of the movement behavior and the quality issues. Therefore, SOMs are applied on the provided data which includes the modeling of the input data to create a useful SOM and to be able to interpret the created SOM and component planes of the SOM. The training process of the SOM is executed using the SOMatic Training application. Thereby, a two-phase training approach is applied decreasing a training error and focusing the trained SOM and more local patterns than on global or regional patterns by simple training process. A component plane catalog is created using the SOMatic Viewer application as well as a Processing Java Sketch to explore the created component planes dynamically. To compare the physical, attribute and time a website is created to explore different spaces. One SOM is providing information about the production asset and the frequency of used equipment and a second SOM is analyzing the quality issue and the frequency of the equipment. Drop-down boxes enable the interaction by changing the component planes by a user interaction as well as the projecting of BMUs onto each SOM per se to identify similarities between the attributes standing behind the BMU including the time. In the physical space, the equipment triggering a quality issue is highlighted to identify potential spatial and temporal patterns.

Indoor Navigation Ontology for a Production Line

The carried out research has a great impact on the development of an indoor ontology for a production line, where agents are moving in the indoor space. The indoor ontology considers an affordance based navigation ontology which looks different than current approaches developed by for example Yang and Worboys (2011), as the indoor environment has production line specific entities. Although, this research is broadening the ontology research in indoor environments. Even by focusing on a specific task, which is the agent based movement the design is applicable to mobile behavior studies and in other fields of the indoor environment.
A further impact of this research is the spatio-temporal database including the general storage of the data, the routing capability, visualization and analysis. The spatio-temporal database is one central data pool which can be an advantage for a company as it is combining their attributive and temporal data which is already stored with a spatial component. Additionally, the data is in a format where it is possible to carry out an analysis and to gain new knowledge out of a standard relational database by being able to visualize the data as a map and analyze it.

To sum up, a new way of visualizing and exploring production line data is applied. This includes the visualization and storage of the data in a spatio-temporal database, the routing capability and comparison of different movement paths as well as the analysis of attribute space, physical space and time. It is an important step, to show new capabilities with the help of GIS and also to gain new knowledge out of production line data to show potential improvements decreasing in example the overall time, the overall distance or to increase the quality of production assets.

8.2 Future Perspectives

The current approach for the spatio-temporal data mining of production line processes shows the opportunity of how to set up an indoor environment of a production line with a GIS and also how to prepare the data to enable a comparison of different movement paths in the production line. This approach demonstrates also the opportunity how it is possible to analyze the movement data and the quality data using a spatio-temporal data technique to gain new knowledge about the current workflow and behavior in the production line.

Future Fields of Application

Until now, the movement paths are extract and analyzed based on historical data. One future issue can be the development of a real-time visualization and monitoring application. Such a real-time application can be linked with historical identified patterns to generate some kind of prediction model. For example, if a certain pattern of equipment can be identified in the attribute space and this pattern is evolving in a currently transported or processed production asset and a warning can be generated. Furthermore, also a spatial pattern can be useful to generate some sort of geo-fencing alert to omit contaminated areas or an example can be to ensure that a production asset with a very high importance is never used by equipment in a certain room resulting such an identified spatial pattern.

Data Preparation and Migration

As it is important to keep the spatial data such as the production line environment and the network up to date by a from time to time changing environment, another future issue is the preparation and migration of the data. A semi-automatic workflow is described in section 5.2 and section 5.3, which is still including high manual effort. Therefore, a future perspective has to consider the development of a fully automatic network generation and preparation of the input file, which is at the moment an AutoCAD file or even a workflow from an already created shapefile to a network and ready to use equipment for the analysis. This is an important issue as it is very time consuming and hard manual work.

Analysis of the Movement Behavior

To enhance the analysis of the movement behavior and the quality issues one opportunity is described in detail and also demonstrated by the implemented website. In addition to this approach, it would be possible to generate different types of input data to probably answer different questions. This can be useful, if the number of input data is decreased. As it is already described in the discussion chapter 7, the visualization of the high amount of production assets or quality issues is a critical issue. On one hand it is important to have a huge amount of data to deliver an accurate analysis, on the other hand it is a complicated task to show an appropriate visualization method. The robustness of carrying out this approach is given

according to the input data. As already mentioned, the selection of a proper class of production assets is recommended as it is simplifying the visual identification of patterns.
The number of quality issues or production assets concerns the computational performance for either the training of a SOM, the general data handling or even the visualization with or without the projecting of BMUs. Therefore, a deeper knowledge of the production assets and their input data or describing data would be useful to create specific limitations of production assets or even smaller groups of production assets. Especially, as the study area is one selected production hall of a whole manufacturing site, which will increase the number of production assets and equipment dramatically if a higher number of production halls will be used. Nevertheless, this approach shows the possibility of comparing the physical space, attribute space and time.

Connection of Indoor and Outdoor Geography

Future research directions can also included the connection between indoor and outdoor space or different types of indoor space as it is already mentioned by Yang and Worboys (2011). In order to evaluate the indoor navigation ontology the identified navigation and movement patterns can be considered. In this approach, the spatio-temporal data mining method SOM is applied, whereas other trajectory pattern mining methods or spatio-temporal data mining methods can be executed.
Patterns in the physical space and the attribute are validated by considering randomly selected production assets. As for example, if production assets have the same recipe or even the same group there is also a certain similarity within the attribute space that can be explored on the SOM.
Another future research direction can be the analysis of the physical space and quality issues per equipment by implementing a dynamic visualization tool. Thereby, a time-range can be defined and moved over the whole provided dataset. This would give an insight and new knowledge about quality affecting specific equipments and in some cases patterns of rooms or whole areas.

Finally, the carried out approach shows the possibility of a spatio-temporal analysis for pattern recognition in a production line as well as there are many further research topics. These research topics concern the increasing field of indoor geography, the development and handling of spatio-temporal data and the overall analysis of movement behavior and quality issues.

Chapter 9

Bibliography

Abraham, S. and Sojan Lal, P. (2012), 'Spatio-temporal similarity of network-constrained moving object trajectories using sequence alignment of travel locations', *Transportation research part C: emerging technologies* **23**, 109–123.

Agarwal, P. and Skupin, A. (2008), *Self-organising maps: Applications in geographic information science*, John Wiley & Sons.

Al Nuaimi, K. and Kamel, H. (2011), A survey of indoor positioning systems and algorithms, *in* 'Innovations in Information Technology (IIT), 2011 International Conference on', IEEE, pp. 185–190.

Alvares, L. O., Bogorny, V., de Macedo, J. A. F., Moelans, B. and Spaccapietra, S. (2007), Dynamic modeling of trajectory patterns using data mining and reverse engineering, *in* 'Tutorials, posters, panels and industrial contributions at the 26th international conference on Conceptual modeling-Volume 83', Australian Computer Society, Inc., pp. 149–154.

Alvares, L., Oliveira, G. and Bogorny, V. (2009), 'A framework for trajectory data preprocessing for data mining', pp. 1 – 5.

Anagnostopoulos, C., Tsetsos, V., Kikiras, P. et al. (2005), Ontonav: A semantic indoor navigation system, *in* '1ST WORKSHOP ON SEMANTICS IN MOBILE ENVIRONMENTS (SME05), AYIA', Citeseer.

Andrienko, G., Andrienko, N., Bremm, S., Schreck, T., Von Landesberger, T., Bak, P. and Keim, D. (2010), Space-in-time and time-in-space self-organizing maps for exploring spatiotemporal patterns, *in* 'Computer Graphics Forum', Vol. 29, Wiley Online Library, pp. 913–922.

Ascraft, H. (2008), 'Building information modeling: A framework for collaboration', *Construction Lawyer* **28**(3), 1 – 14.

Barnes, J., Rizos, C., Wang, J., Small, D., Voigt, G. and Gambale, N. (2003), 'High precision indoor and outdoor positioning using locatanet', *Journal of Global Positioning System* **2**(2), 73–82.

Becker, T., Nagel, C. and Kolbe, T. H. (2009), A multilayered space–event model for navigation in indoor spaces, *in* '3D Geo-Information Sciences', Springer, pp. 61–77.

Ben Aoun, N., Mejdoub, M. and Ben Amar, C. (2014), 'Graph-based approach for human action recognition using spatio-temporal features', *Journal of Visual Communication and Image Representation* **25**(2), 329–338.

Bill, R., Cap, C., Kofahl, M. and Mundt, T. (2004), 'Indoor and outdoor positioning in mobile environments : a review and some investigations on wlan-positioning', *Geographic Information Sciences* **10**(2).

Bogorny, V., Palma, A., Engel, P. and Alvares, L. (2006), 'Weka-gdpm-integrating classical data mining toolkit to gegraphic information systems', pp. 1 – 8.

Compieta, P., Di Martino, S., Bertolotto, M., Ferrucci, F. and Kechadi, T. (2007), 'Exploratory spatio-temporal data mining and visualization', *Journal of Visual Languages & Computing* **18**(3), 255–279.

Davis, E. (1990), *Representations of commonsense knowledge*, Morgan Kaufmann Publishers Inc.

Erwig, M., Gu, R. H., Schneider, M., Vazirgiannis, M. et al. (1999), 'Spatio-temporal data types: An approach to modeling and querying moving objects in databases', *GeoInformatica* **3**(3), 269–296.

Fan, Y., Yang, J., Zhu, D. and Wei, K. (2010), 'A time-based integration method of spatio-temporal data at spatial database level', *Mathematical and Computer Modelling* **51**(11), 1286–1292.

Fan, Y., Yang, J., Zhu, D. and Zhang, C. (2011), 'Research on the efficiency of querying historical data with the spatio-time data integration method', *Mathematical and Computer Modelling* **54**(3), 912–918.

Gibson, J. (1977), 'The concept of affordances', *Perceiving, acting, and knowing* pp. 67–82.

Giudice, N. A., Walton, L. A. and Worboys, M. (2010), The informatics of indoor and outdoor space: a research agenda, *in* 'Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness', ACM, pp. 47–53.

Goetz, M. (2012), 'Using crowdsourced indoor geodata for the creation of a three-dimensional indoor routing web application', *Future Internet* **4**(2), 575–591.

Goetz, M. and Zipf, A. (2011), 'Formal definition of a user–adaptive and length–optimal routing graph for complex indoor environments', *Geo-Spatial Information Science* **14**(2), 119–128.

Hagedorn, B., Trapp, M., Glander, T. and Dollner, J. (2009), Towards an indoor level–of–detail model for route visualization, *in* 'Mobile Data Management: Systems, Services and Middleware, 2009. MDM'09. Tenth International Conference on', IEEE, pp. 692–697.

Howell, I. and Batcheler, B. (2005), 'Building information modeling two years later–huge potential, some success and several limitations', *The Laiserin Letter* **22**.

Isikdag, U., Zlatanova, S. and Underwood, J. (2013), 'A bim-oriented model for supporting indoor navigation requirements', *Computers, Environment and Urban Systems* **41**, 112–123.

Jensen, C. S., Lu, H. and Yang, B. (2009), Graph model based indoor tracking, *in* 'Mobile Data Management: Systems, Services and Middleware, 2009. MDM'09. Tenth International Conference on', IEEE, pp. 122–131.

Jeung, H., Yiu, M. L. and Jensen, C. S. (2011), Trajectory pattern mining, *in* 'Computing with Spatial Trajectories', Springer, pp. 143–177.

Kapadia, M., Singh, S., Hewlett, W. and Faloutsos, P. (2009), Egocentric affordance fields in pedestrian steering, *in* 'Proceedings of the 2009 symposium on Interactive 3D graphics and games', ACM, pp. 215–223.

Kim, Y. S., Lee, S. W., Kim, S. R., Jeong, H. and Kim, J. H. (2012), A product-service systems design method with integration of product elements and service elements using affordances, *in* 'ServDes–3rd Service Design and Service Innovation Conference, Helsinki'.

Kohonen, T. (1998), 'The self-organizing map', *Neurocomputing* **21**(1), 1–6.

Kohonen, T. (2012), 'Essentials of the self-organizing map', *Neural Networks* .

Kohonen, T., Hynninen, J., Kangas, J. and Laaksonen, J. (1996), 'Som pak: The self-organizing map program package', *Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science* .

Koyuncu, H. and Yang, S. H. (2010), 'A survey of indoor positioning and object locating systems', *IJCSNS International Journal of Computer Science and Network Security* **10**(5), 121–128.

Lee, A. J., Chen, Y.-A. and Ip, W.-C. (2009), 'Mining frequent trajectory patterns in spatial–temporal databases', *Information Sciences* **179**(13), 2218–2231.

Li, B., Quader, I. J. and Dempster, A. G. (2008), 'On outdoor positioning with wi-fi', *Journal of Global Positioning Systems* **7**(1), 18–26.

Mamoulis, N., Cao, H., Kollios, G., Hadjieleftheriou, M., Tao, Y. and Cheung, D. W. (2004), Mining, indexing, and querying historical spatiotemporal data, *in* 'Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining', ACM, pp. 236–245.

Mautz, R. (2008), Combination of indoor and outdoor positioning, *in* '1st International Conference on Machine Control & Guidance'.

Mautz, R. (2009), 'Overview of current indoor positioning systems', *Geodezija ir kartografija* **35**(1), 18–22.

Meijers, M., Zlatanova, S. and Pfeifer, N. (2005), 3d geoinformation indoors: structuring for evacuation, *in* 'Proceedings of Next generation 3D city models', pp. 21–22.

Nyström, R. H., Harjunkoski, I. and Kroll, A. (2006), 'Production optimization for continuously operated processes with optimal operation and scheduling of multiple units', *Computers & chemical engineering* **30**(3), 392–406.

Papadias, D., Tao, Y., Kanis, P. and Zhang, J. (2002), Indexing spatio-temporal data warehouses, *in* 'Data Engineering, 2002. Proceedings. 18th International Conference on', IEEE, pp. 166–175.

Pelekis, N., Theodoulidis, B., Kopanakis, I. and Theodoridis, Y. (2004), 'Literature review of spatio-temporal database models', *The Knowledge Engineering Review* **19**(03), 235–274.

Rainer, M. (2013), Somatic viewer: Implementation of an interactive self-organizing map visualization toolset in processing and java, Master's thesis, Spatial Information Management, Carinthia University of Applied Sciences.

Raubal, M. (2001), 'Ontology and epistemology for agent-based wayfinding simulation', *International Journal of Geographical Information Science* **15**(7), 653–665.

Raubal, M. and Worboys, M. (1999), A formal model of the process of wayfinding in built environments, *in* 'Spatial information theory. Cognitive and computational foundations of geographic information science', Springer, pp. 381–399.

Schneider, M. (2009), 'Spatial and spatio-temporal data models and languages.'.

Scholz, J. and Schabus, S. (2014), An indoor navigation ontology for production assets in a production environment, *in* 'Proceedings of GIScience 2014, Vienna, Austria - Lecture Notes in Computer Science, Springer'.

Skupin, A. (2010), Tri-space: Conceptualization, transformation, visualization, *in* 'Sixth International Conference on Geographic Information Science', pp. 14–17.

Skupin, A. and Esperbé, A. (2011), 'An alternative map of the united states based on an n-dimensional model of geographic space', *Journal of Visual Languages & Computing* **22**(4), 290–304.

Smith, A., Balakrishnan, H., Goraczko, M. and Priyantha, N. (2004), Tracking moving devices with the cricket location system, *in* 'Proceedings of the 2nd international conference on Mobile systems, applications, and services', ACM, pp. 190–202.

Smith, B. (2001), 'Objects and their environments: from aristotle to ecological ontology', *Life and Motion of Socio-economic Units. GISDATA Series* **8**, 79–97.

Spöcklberger, M. (2013), Somatic trainer: Implementation of a self-organizing map tool with parallel training for processing, Master's thesis, Spatial Information Management, Carinthia University of Applied Sciences.

Stoffel, E.-P., Schoder, K. and Ohlbach, H. J. (2008), Applying hierarchical graphs to pedestrian indoor navigation, *in* 'Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems', ACM, p. 54.

Theiss, A., Yen, D. C. and Ku, C.-Y. (2005), 'Global positioning systems: an analysis of applications, current development and future implementations', *Computer Standards & Interfaces* **27**(2), 89–100.

Thiesse, F., Fleisch, E. and Dierkes, M. (2006), 'Lottrack: Rfid-based process control in the semiconductor industry', *Pervasive Computing, IEEE* **5**(1), 47–53.

Tsetsos, V., Anagnostopoulos, C., Kikiras, P. and Hadjiefthymiades, S. (2006), 'Semantically enriched navigation for indoor environments', *International Journal of Web and Grid Services* **2**(4), 453–478.

Turner, A. and Penn, A. (2002), 'Encoding natural movement as an agent-based system: an investigation into human pedestrian behaviour in the built environment', *Environ Plann B* **29**(4), 473–490.

Vesanto, J., Himberg, J., Alhoniemi, E. and Parhankangas, J. (2000), *SOM toolbox for Matlab 5*, Citeseer.

Volk, R., Stengel, J. and Schultmann, F. (2014), 'Building information modeling (bim) for existing buildings–literature review and future needs', *Automation in Construction* **38**, 109–127.

Worboys, M. (2011), Modeling indoor space, *in* 'Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness', ACM, pp. 1–6.

Xiang, Z., Song, S., Chen, J., Wang, H., Huang, J. and Gao, X. (2004), 'A wireless lan-based indoor positioning technology', *IBM Journal of Research and Development* **48**(5.6), 617–626.

Yang, L. and Worboys, M. (2011), A navigation ontology for outdoor-indoor space:(work-in-progress), *in* 'Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness', ACM, pp. 31–34.

Yeh, S.-C., Hsu, W.-H., Su, M.-Y., Chen, C.-H. and Liu, K.-H. (2009), A study on outdoor positioning technology using gps and wifi networks, *in* 'Networking, Sensing and Control, 2009. ICNSC'09. International Conference on', IEEE, pp. 597–601.

List of Tables

List of Figures

Appendix A

Use Cases of the Production Line System

Definition of the use cases from sub-section 4.4.1 in a tabular form. The use cases are described including their actor and a detailed description of the use case. Further, the pre-condition, the post-condition and the flow of events are stated for the specific use case.

Table A.1: Use case specification - Take Production Asset.

Use Case 1 - *Take Production Asset*	
Initiating Actor:	Transport Operator
Actor Goal:	Grab a production asset and decide next processing steps.
Pre-Condition:	– Production Asset finished processing at an equipment – Production Asset is in a shelf and has to be transported.
Post-Condition:	– Production Asset is waiting to be processed by an equipment. – Production Asset is in a shelf for a further transport.
Flow of events:	– Transport Operator selects a Production Asset. – Checks the Production Assets priority. – Identifies the next transportation target.

Table A.2: Use case specification - Take Transportation Cart.

Use Case 2 - *Take Transportation Cart*	
Initiating Actor:	Transport Operator
Actor Goal:	Take a cart to transport several production assets at once.
Pre-Condition:	– Empty Cart before the transport – Several Production Assets are ready.
Post-Condition:	– Transportation cart is full with production assets.
Flow of events:	– Take a transportation cart. – Includes *use case 1*, take production asset.

Table A.3: Use case specification - Transporting Activity.

Use Case 3 - *Transporting Activity*	
Initiating Actor:	Transport Operator
Actor Goal:	The actor is transporting the production assets to the identified equipment or facility.
Pre-Condition:	– Single Production Asset has to be taken. – Transportation cart full with production assets.
Post-Condition:	– Single Production Asset delivered to next shelf or equipment. – All Production Assets from the cart are delivered.
Flow of events:	– Select a production asset or many assets on a cart. – Deliver all production assets to the next equipment, shelf. – Return with empty cart.

Table A.4: Use case specification - Communicating Information.

Use Case 4 - *Communicating Information*	
Initiating Actor:	Positioning Solution
Actor Goal:	Provide the transport operator with useful information about the production asset (Next equipment, next facility ,priority ,...).
Pre-Condition:	– Production Asset has to be equipped with a tracking and communicating device.
Post-Condition:	– Transport Operator sees transmitted information. – Information has to be provided.
Flow of events:	– Transport Operator has to check information on the display of the communication device.

Table A.5: Use case specification - Tracking Production Asset.

Use Case 5 - *Tracking Production Asset*	
Initiating Actor:	Positioning Solution
Actor Goal:	Track the production asset in the production line the whole time.
Pre-Condition:	– Production Asset has to be equipped with a tracking and communicating device. – working positioning solution.
Post-Condition:	– Production Asset is tracked throughout the processes. – Position data is stored.
Flow of events:	– Production Asset is equipped with tracking device. – Implemented Positioning Solution – Includes storing the position information in a database.

Table A.6: Use case specification - Processing Production Asset.

Use Case 6 - *Processing Production Asset*	
Initiating Actor:	Equipment
Actor Goal:	Active processing of a production asset as one step of many.
Pre-Condition:	– Equipment has to be ready for processing. – Equipment has to support the specific operation.
Post-Condition:	– Finished processing of Production Asset. – Ready for transport to next operation.
Flow of events:	– Production Asset is checking in at equipment. – Processing of the production asset. – Checking out and transmitting information.

Table A.7: Use case specification - Analyze Activities of Production Asset.

Use Case 7 - *Analyze Activities of Production Asset*	
Initiating Actor:	Control Operator
Actor Goal:	Check carried out activities and stored information for quality tasks.
Pre-Condition:	– Production Asset is selected for analysis. – Data has to be present in the database system.
Post-Condition:	– Quality check of operations. – Information about the movement behavior.
Flow of events:	– Select Production Asset and gather data. – Start analysis. – Make decisions for a proper development.

Table A.8: Use case specification - Checking in at Equipment.

Use Case 8 - *Checking in at Equipment*	
Initiating Actor:	Production Asset
Actor Goal:	Check in at equipment for further processing.
Pre-Condition:	– Production Asset is ready for being processed.
Post-Condition:	– Production Asset is being processed.
Flow of events:	– Operator puts Production Asset in an equipment. – Production Assets checks in and proofs operation. – Stores information in the database system.

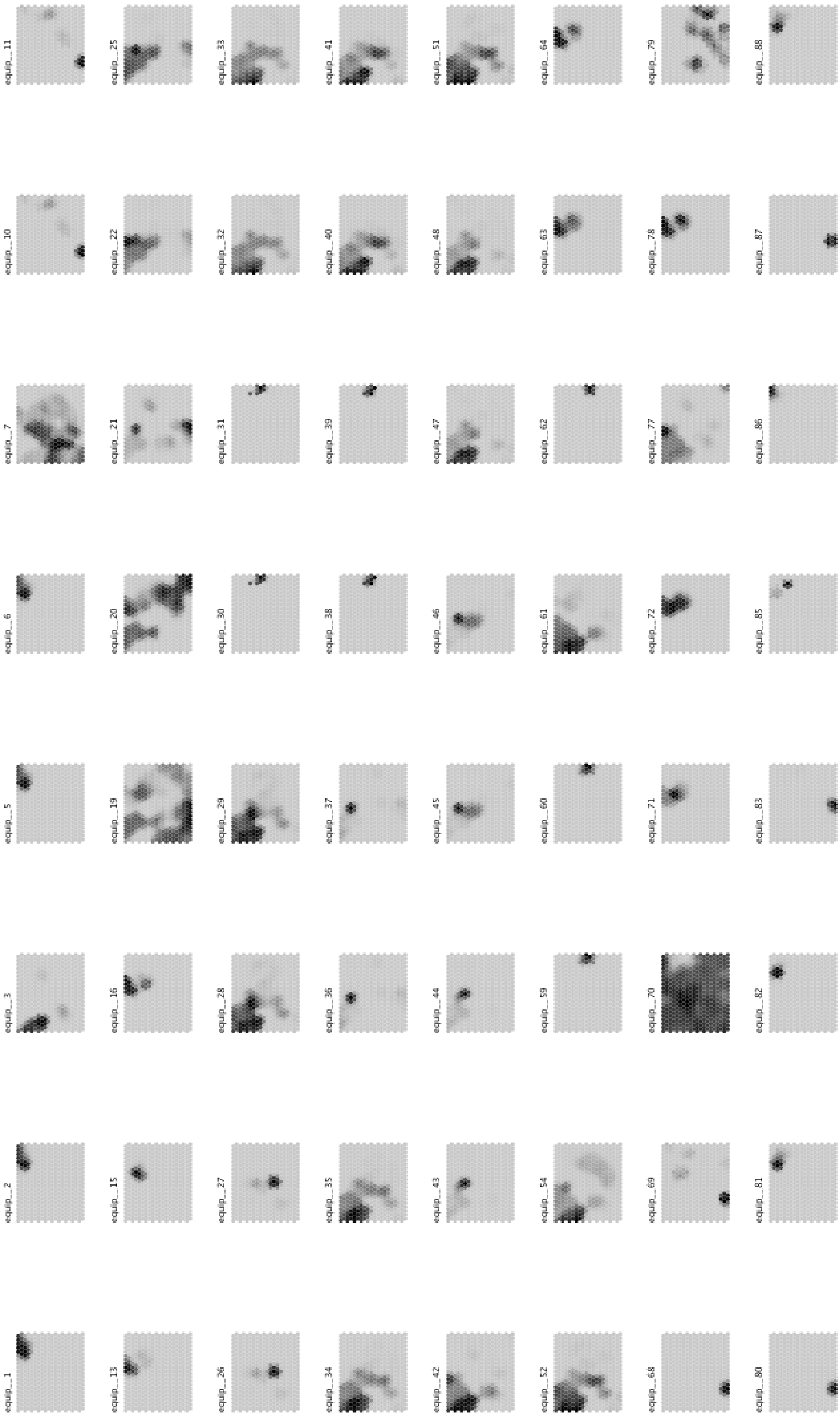
Table A.9: Use case specification - Checking out at Equipment.

Use Case 9 - *Checking out at Equipment*	
Initiating Actor:	Production Asset
Actor Goal:	Check out at equipment after successfully processing.
Pre-Condition:	– Production Asset finished processing at equipment.
Post-Condition:	– Production Asset is ready for transport and next operation.
Flow of events:	– Equipment finishes processing. – Production Assets checks out. – Stores information in the database system.

Appendix B

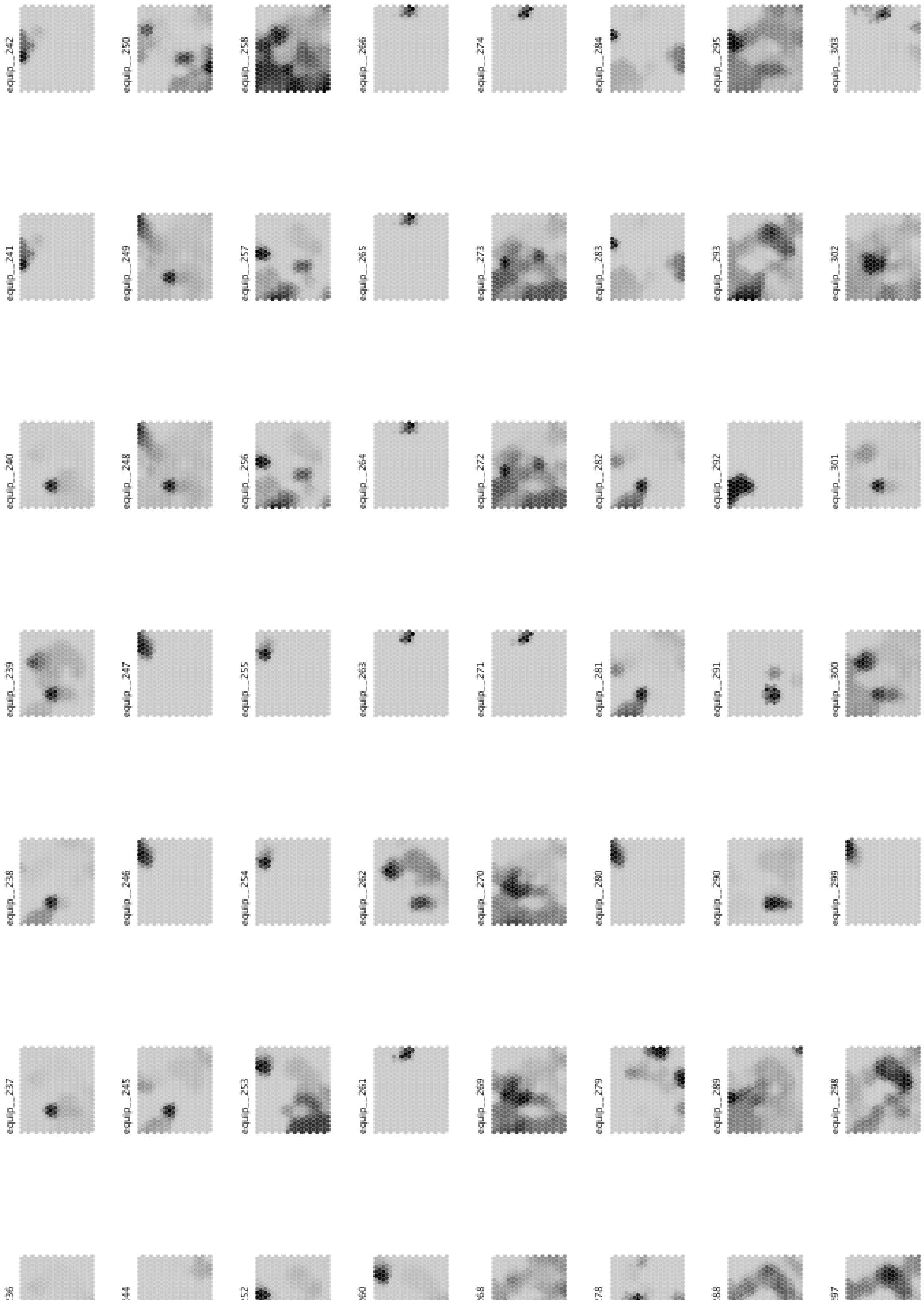
Component Planes - Production Asset and Equipment

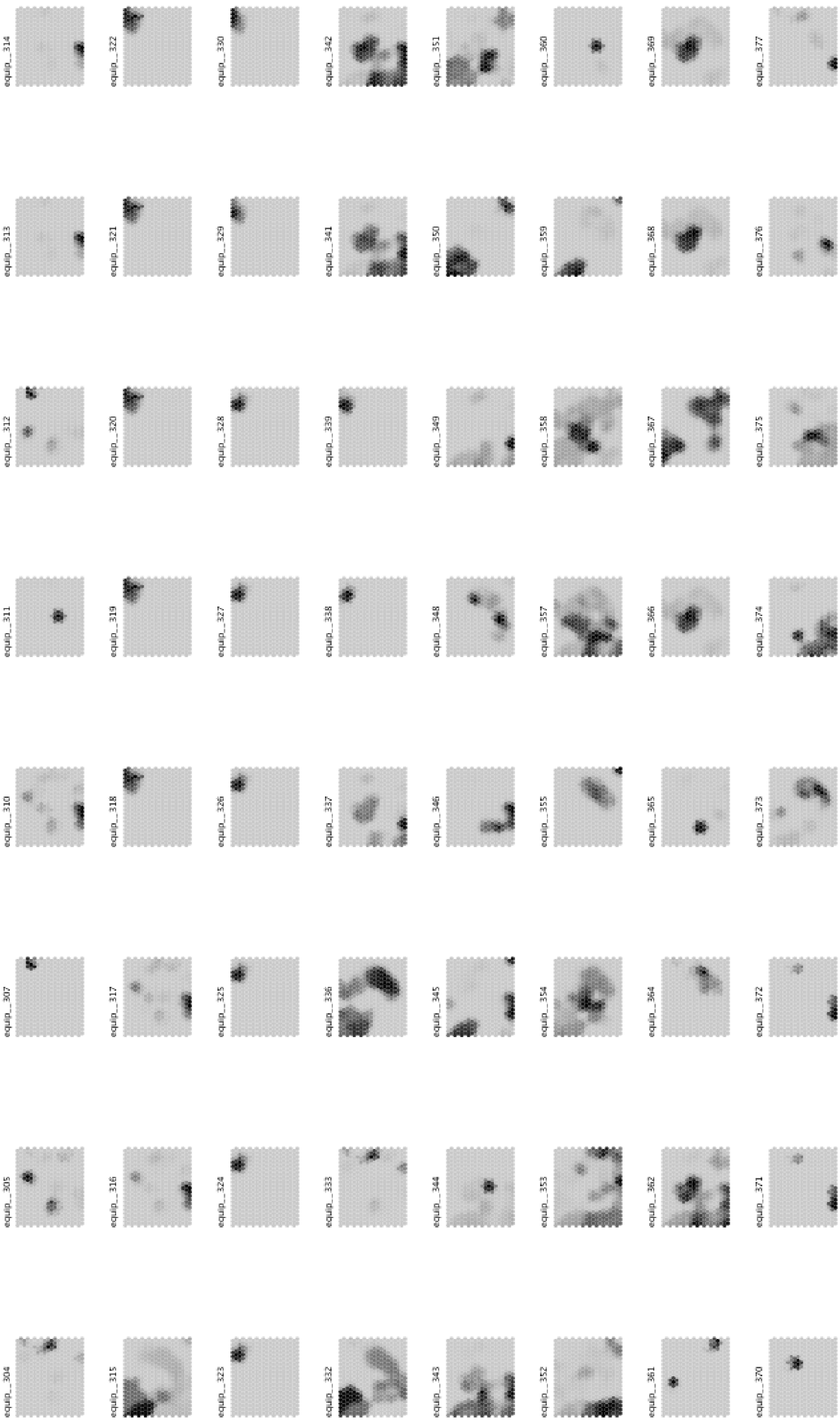
Appendix B is providing a Catalog of all the created Component Planes of the SOM covering the frequency of the equipment per each production asset. The SOM is trained using the SOMatic Training Application with the training parameters described in section 5.6.1. The SOM is visualized with the SOMatic Viewer Application described in section 5.6.2. In the result section 6.5 for the Component Planes of the production asset and the equipment some examples are shown and described. A short overview is given how it is possible to interpret the corresponding component planes which are all listed in this component plane catalog. All equipments are listed with the following abbreviation "equip" for equipment and the number is the identifier for the equipment in the spatio-temporal database. The equipment component planes are ordered ascending by their identifier. Some equipments or in this case number does not appear in the component plane catalog as they are never frequented by an production asset and the component plane would be trivial. The component planes can be interpreted by considering the color schema where a dark value stands for a high frequency and the opposite are low values or low frequencies. If different component planes have neurons with the same color or that are very close to each other they are more similar, than neurons that are further apart. Some component planes have high values and are distributed and others are focused in specific and small areas. This leads to possible interpretations about the production asset and the used equipment.

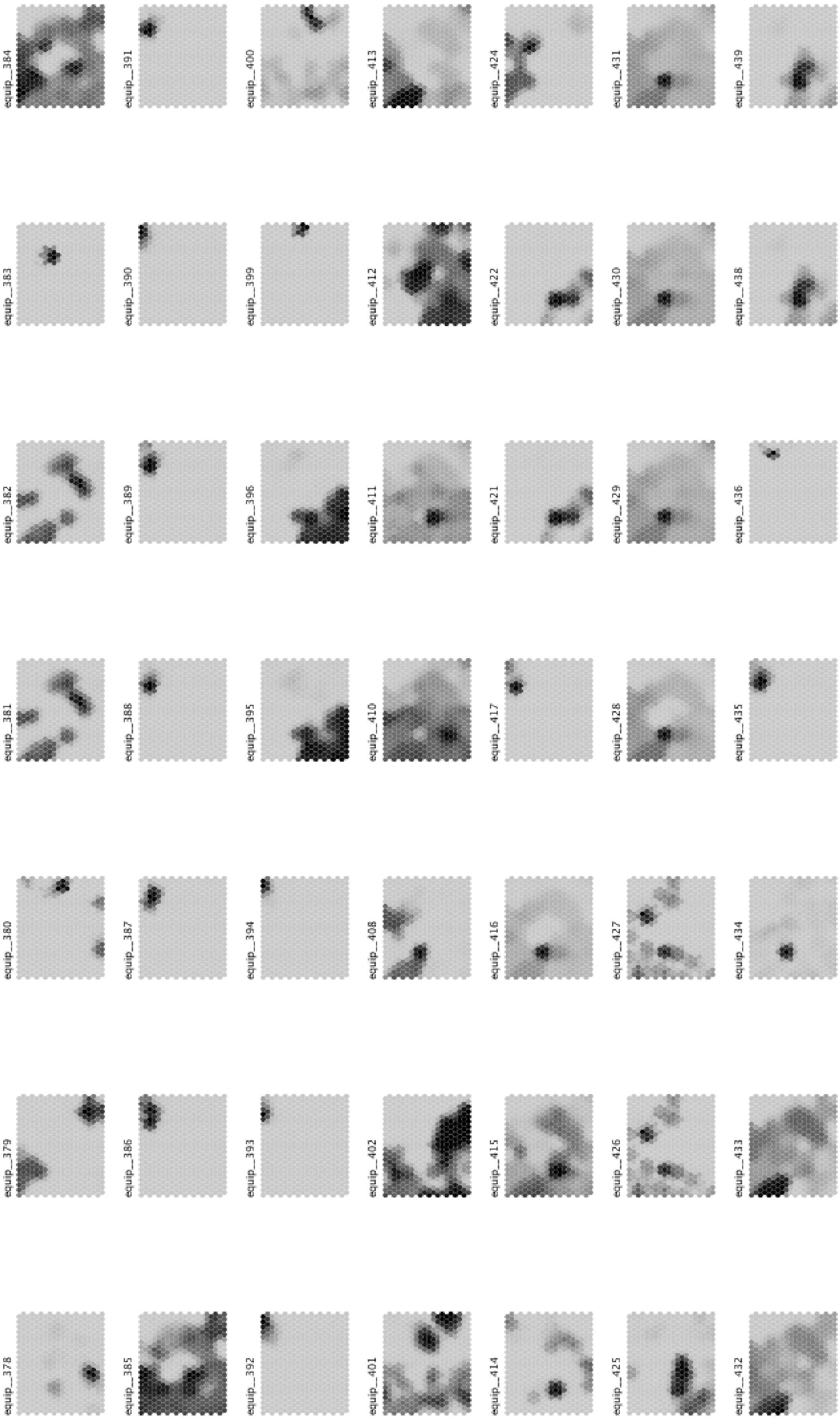

equip__1
equip__2
equip__3
equip__5
equip__6
equip__7
equip__10
equip__11
equip__13
equip__15
equip__16
equip__19
equip__20
equip__21
equip__22
equip__25
equip__26
equip__27
equip__28
equip__29
equip__30
equip__31
equip__32
equip__33
equip__34
equip__35
equip__36
equip__37
equip__38
equip__39
equip__40
equip__41
equip__42
equip__43
equip__44
equip__45
equip__46
equip__47
equip__48
equip__51
equip__52
equip__54
equip__59
equip__60
equip__61
equip__62
equip__63
equip__64
equip__68
equip__69
equip__70
equip__71
equip__72
equip__77
equip__78
equip__79
equip__80
equip__81
equip__82
equip__83
equip__85
equip__86
equip__87
equip__88

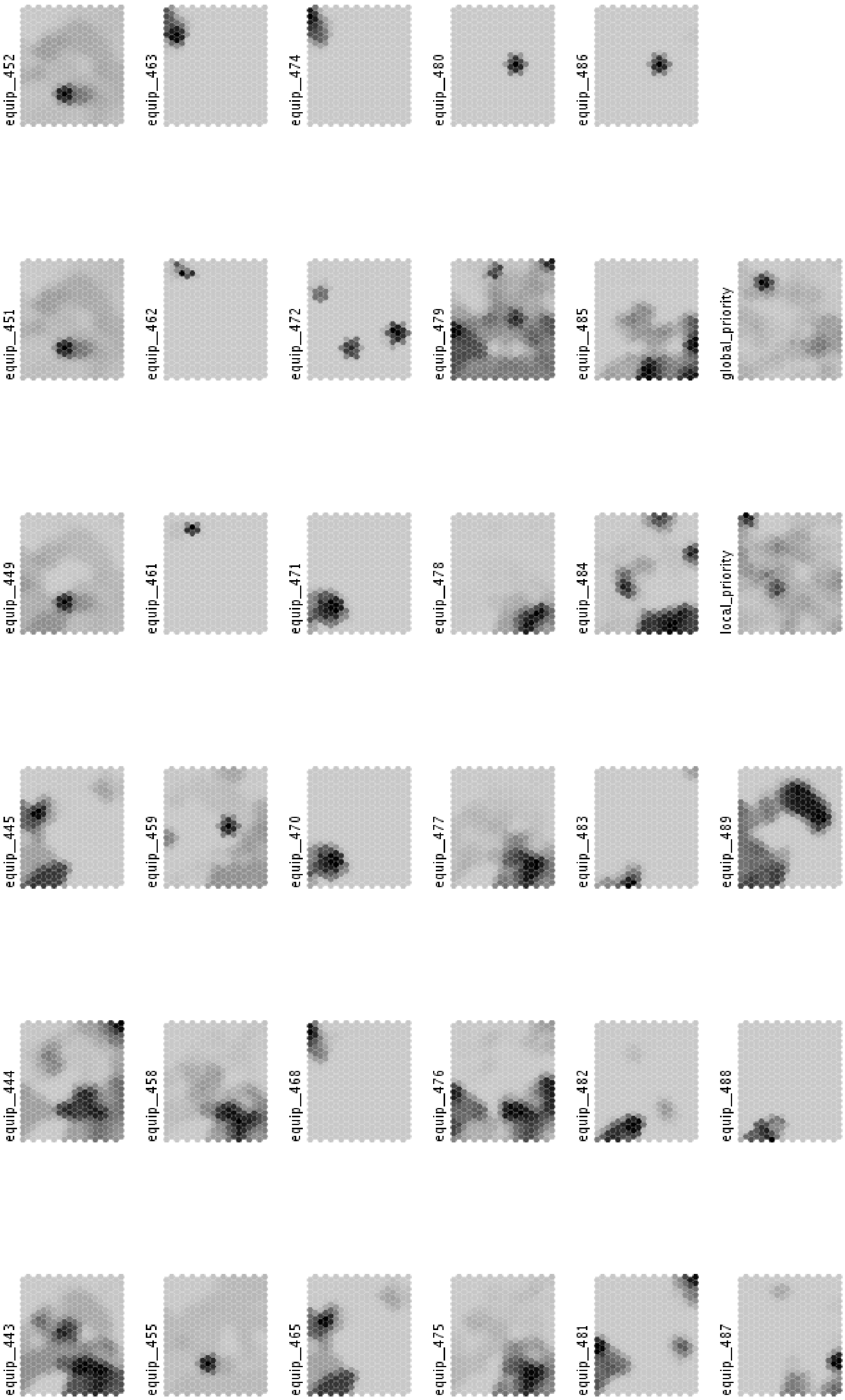

equip__89
equip__90
equip__91
equip__94
equip__95
equip__96
equip__97
equip__98
equip__99
equip__100
equip__101
equip__102
equip__103
equip__104
equip__105
equip__106
equip__107
equip__108
equip__109
equip__110
equip__113
equip__114
equip__115
equip__116
equip__117
equip__118
equip__119
equip__120
equip__121
equip__122
equip__123
equip__124
equip__125
equip__126
equip__127
equip__128
equip__129
equip__130
equip__131
equip__132
equip__133
equip__134
equip__135
equip__136
equip__137
equip__138
equip__139
equip__140
equip__141
equip__142
equip__143
equip__144
equip__145
equip__146
equip__147
equip__148
equip__149
equip__150
equip__151
equip__152
equip__153
equip__154
equip__155
equip__156


equip_157
equip_159
equip_160
equip_161
equip_162
equip_163
equip_164
equip_165
equip_166
equip_167
equip_168
equip_171
equip_172
equip_173
equip_174
equip_175
equip_176
equip_179
equip_180
equip_181
equip_182
equip_183
equip_184
equip_185
equip_186
equip_187
equip_188
equip_189
equip_190
equip_191
equip_192
equip_193
equip_194
equip_195
equip_196
equip_197
equip_198
equip_199
equip_200
equip_201
equip_202
equip_203
equip_204
equip_205
equip_206
equip_207
equip_208
equip_209
equip_210
equip_211
equip_212
equip_213
equip_214
equip_215
equip_216
equip_218
equip_219
equip_220
equip_221
equip_222
equip_223
equip_224
equip_225
equip_226
equip_227
equip_228
equip_229
equip_230
equip_231
equip_232
equip_233
equip_234


equip__235
equip__236
equip__237
equip__238
equip__239
equip__240
equip__241
equip__242
equip__243
equip__244
equip__245
equip__246
equip__247
equip__248
equip__249
equip__250
equip__251
equip__252
equip__253
equip__254
equip__255
equip__256
equip__257
equip__258
equip__259
equip__260
equip__261
equip__262
equip__263
equip__264
equip__265
equip__266
equip__267
equip__268
equip__269
equip__270
equip__271
equip__272
equip__273
equip__274
equip__275
equip__278
equip__279
equip__280
equip__281
equip__282
equip__283
equip__284
equip__287
equip__288
equip__289
equip__290
equip__291
equip__292
equip__293
equip__295
equip__296
equip__297
equip__298
equip__299
equip__300
equip__301
equip__302
equip__303


equip__304
equip__305
equip__307
equip__310
equip__311
equip__312
equip__313
equip__314
equip__315
equip__316
equip__317
equip__318
equip__319
equip__320
equip__321
equip__322
equip__323
equip__324
equip__325
equip__326
equip__327
equip__328
equip__329
equip__330
equip__332
equip__333
equip__336
equip__337
equip__338
equip__339
equip__341
equip__342
equip__343
equip__344
equip__345
equip__346
equip__348
equip__349
equip__350
equip__351
equip__352
equip__353
equip__354
equip__355
equip__357
equip__358
equip__359
equip__360
equip__361
equip__362
equip__364
equip__365
equip__366
equip__367
equip__368
equip__369
equip__370
equip__371
equip__372
equip__373
equip__374
equip__375
equip__376
equip__377


equip__378
equip__379
equip__380
equip__381
equip__382
equip__383
equip__384
equip__385
equip__386
equip__387
equip__388
equip__389
equip__390
equip__391
equip__392
equip__393
equip__394
equip__395
equip__396
equip__399
equip__400
equip__401
equip__402
equip__408
equip__410
equip__411
equip__412
equip__413
equip__414
equip__415
equip__416
equip__417
equip__421
equip__422
equip__424
equip__425
equip__426
equip__427
equip__428
equip__429
equip__430
equip__431
equip__432
equip__433
equip__434
equip__435
equip__436
equip__438
equip__439

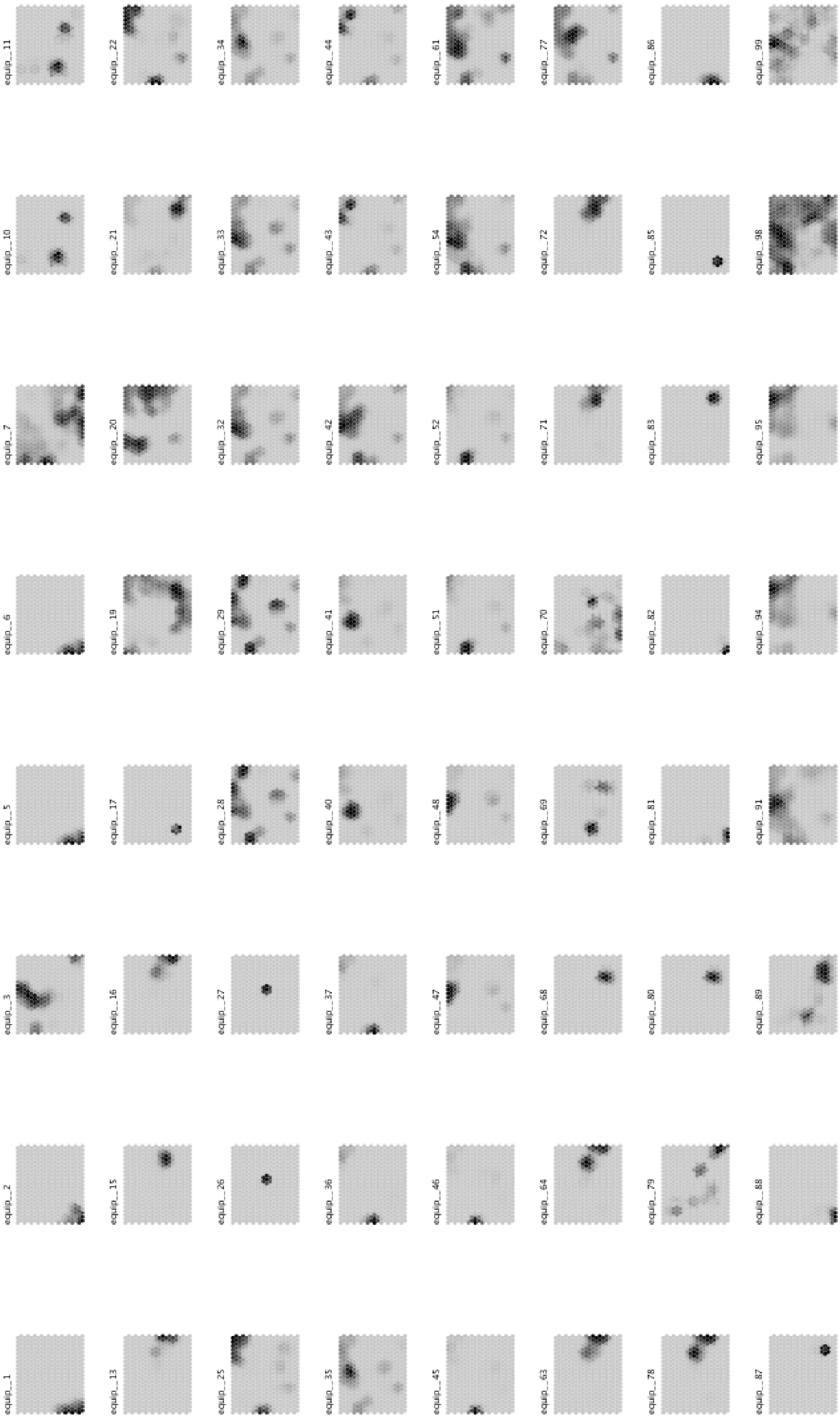

equip__443
equip__444
equip__445
equip__449
equip__451
equip__452
equip__455
equip__458
equip__459
equip__461
equip__462
equip__463
equip__465
equip__468
equip__470
equip__471
equip__472
equip__474
equip__475
equip__476
equip__477
equip__478
equip__479
equip__480
equip__481
equip__482
equip__483
equip__484
equip__485
equip__486
equip__487
equip__488
equip__489
local_priority
global_priority

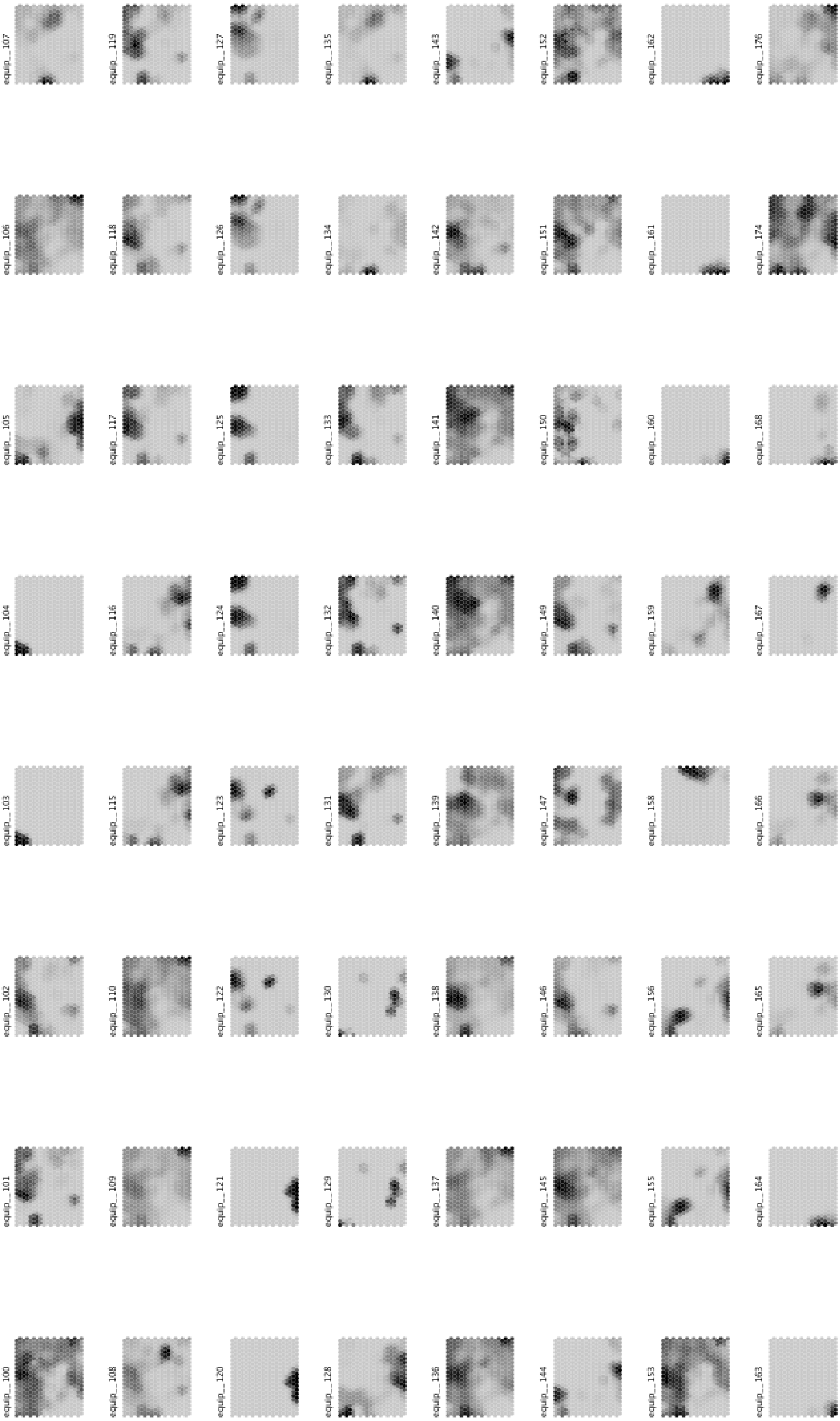
Appendix C

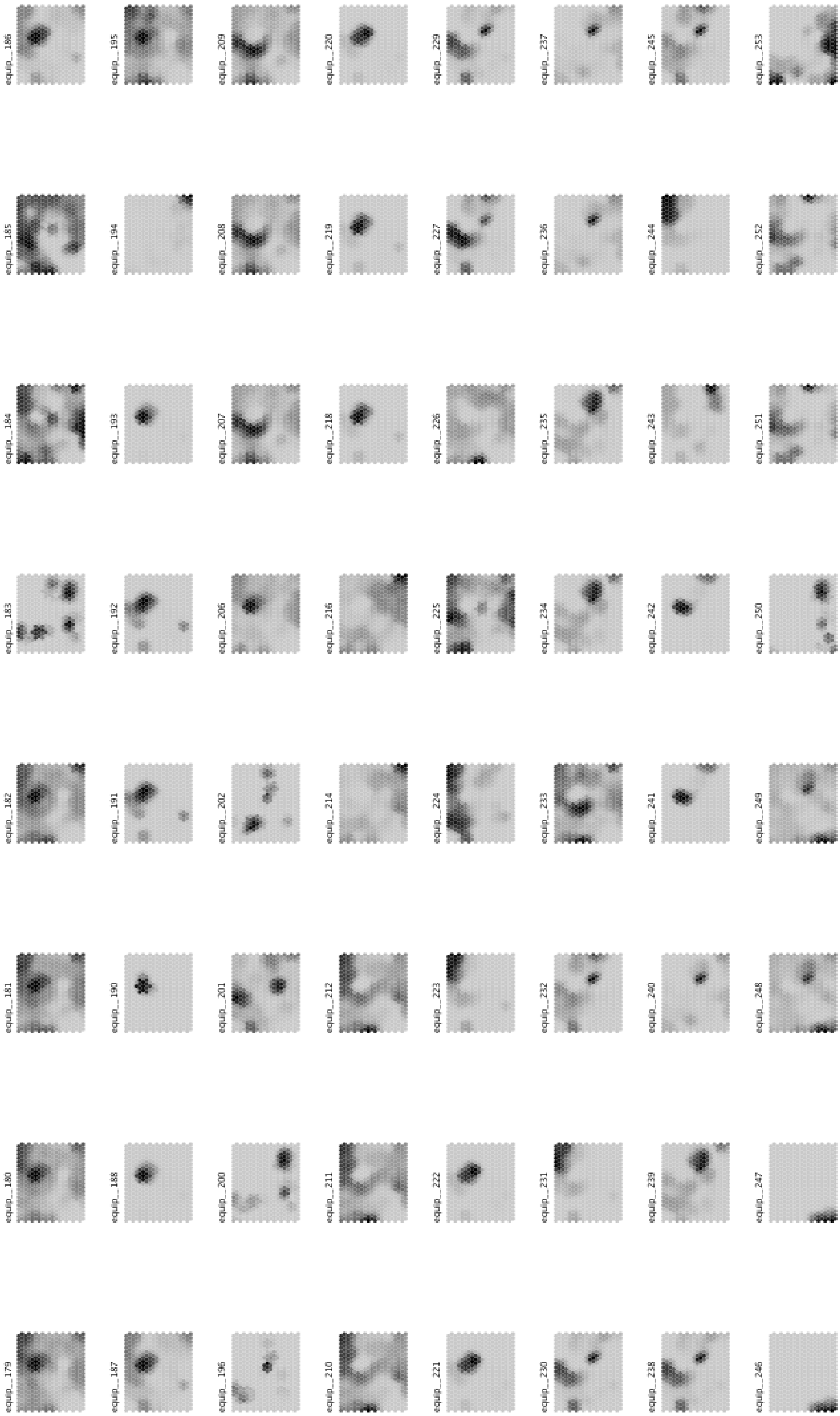
Component Planes - Quality Issue and Equipment

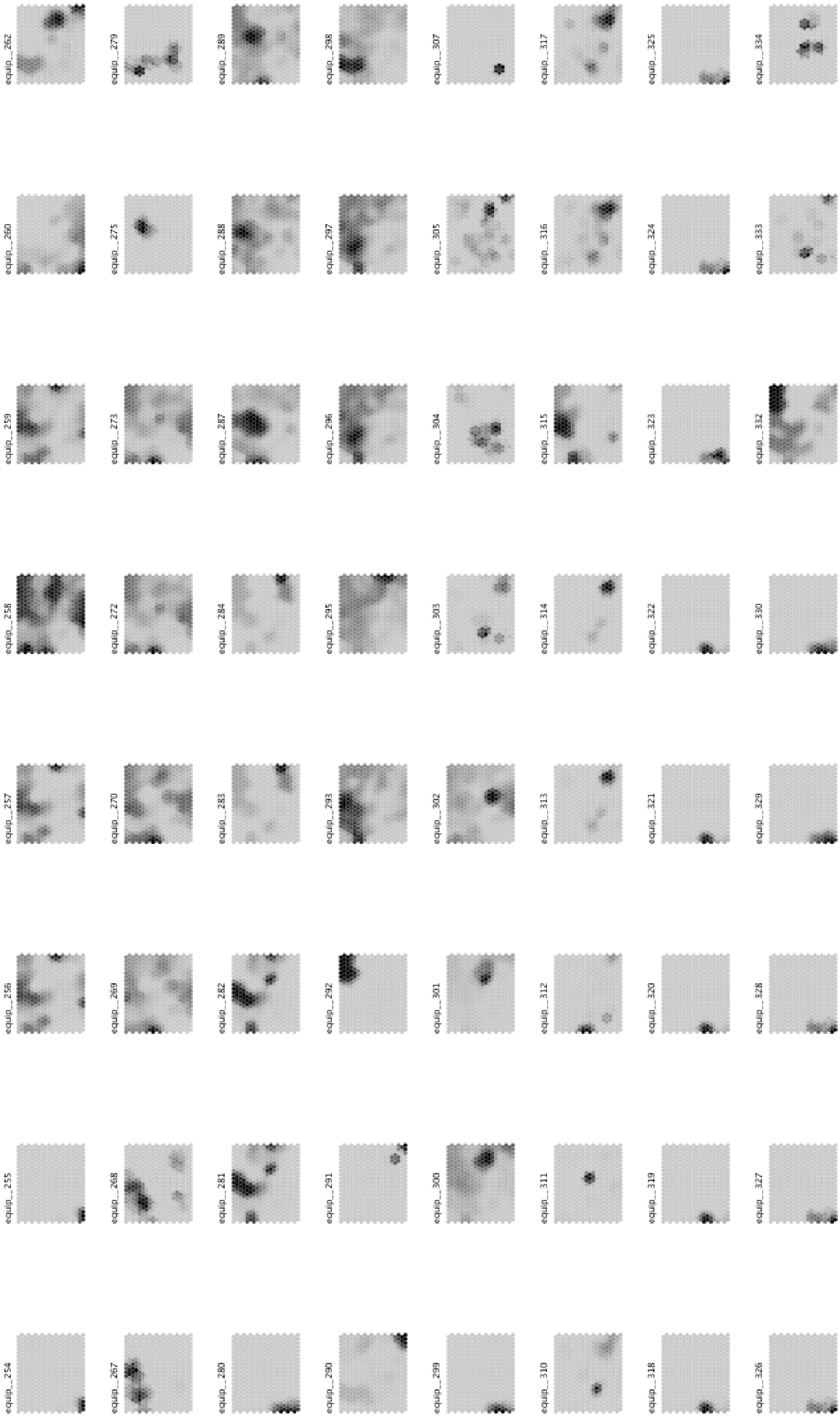
Appendix C is providing a Catalog of all the created Component Planes of the SOM covering the frequency of the equipment per quality issue. The SOM is trained using the SOMatic Training Application with the training parameters described in section 5.6.1. The SOM is visualized with the SOMatic Viewer Application described in section 5.6.2. In the result section 6.6 for the Component Planes of the quality issue and the equipment some examples are shown and described. A short overview is given how it is possible to interpret the corresponding component planes which are all listed in this component plane catalog. All equipments are listed with the following abbreviation "equip" for equipment and the number is the identifier for the equipment in the spatio-temporal database. The equipment component planes are ordered ascending by their identifier. Some equipments or in this case number does not appear in the component plane catalog as they are never frequented by a production asset and the component plane would be trivial.

The component planes can be interpreted by considering the color schema where a dark value stands for a high frequency and the opposite are low values or low frequencies. If different component planes have neurons with the same color or that are very close to each other they are more similar, than neurons that are further apart. Some component planes have high values and are distributed and others are focused in specific and small areas. This leads to possible interpretations about the quality issues and the concerning equipment.

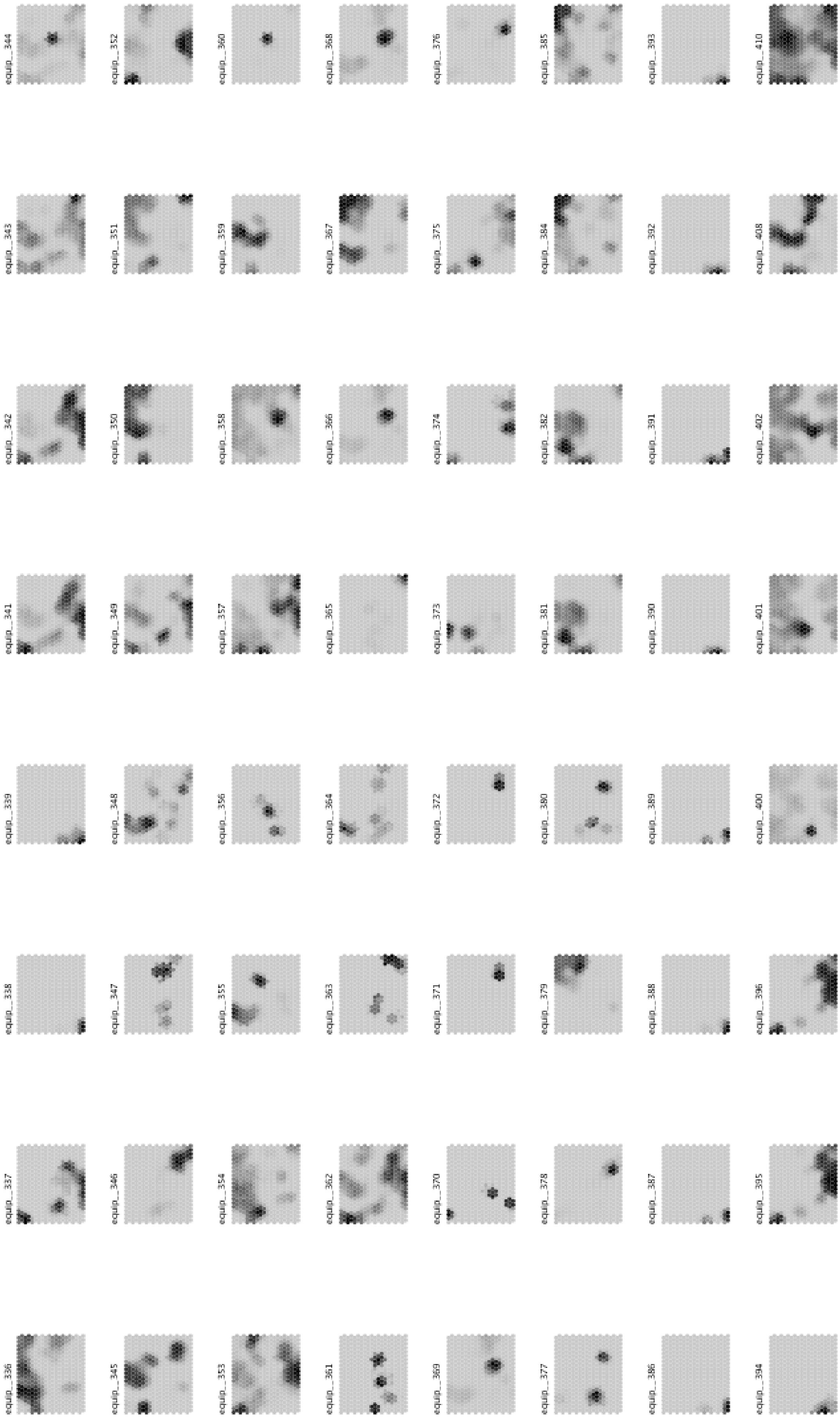

equip__1
equip__2
equip__3
equip__5
equip__6
equip__7
equip__10
equip__11
equip__13
equip__15
equip__16
equip__17
equip__19
equip__20
equip__21
equip__22
equip__25
equip__26
equip__27
equip__28
equip__29
equip__32
equip__33
equip__34
equip__35
equip__36
equip__37
equip__40
equip__41
equip__42
equip__43
equip__44
equip__45
equip__46
equip__47
equip__48
equip__51
equip__52
equip__54
equip__61
equip__63
equip__64
equip__68
equip__69
equip__70
equip__71
equip__72
equip__77
equip__78
equip__79
equip__80
equip__81
equip__82
equip__83
equip__85
equip__86
equip__87
equip__88
equip__89
equip__91
equip__94
equip__95
equip__98
equip__99

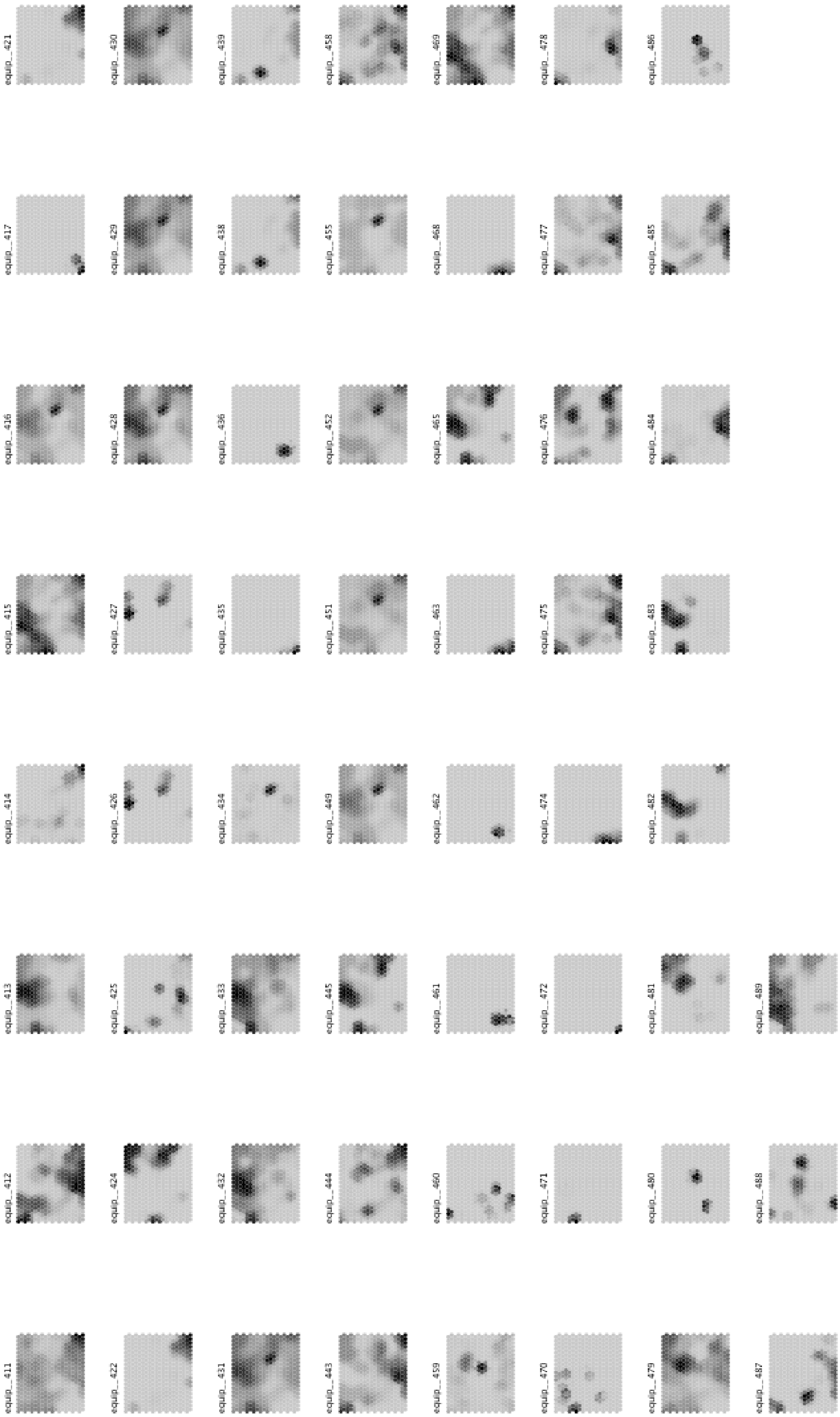

equip__100
equip__101
equip__102
equip__103
equip__104
equip__105
equip__106
equip__107
equip__108
equip__109
equip__110
equip__115
equip__116
equip__117
equip__118
equip__119
equip__120
equip__121
equip__122
equip__123
equip__124
equip__125
equip__126
equip__127
equip__128
equip__129
equip__130
equip__131
equip__132
equip__133
equip__134
equip__135
equip__136
equip__137
equip__138
equip__139
equip__140
equip__141
equip__142
equip__143
equip__144
equip__145
equip__146
equip__147
equip__149
equip__150
equip__151
equip__152
equip__153
equip__155
equip__156
equip__158
equip__159
equip__160
equip__161
equip__162
equip__163
equip__164
equip__165
equip__166
equip__167
equip__168
equip__174
equip__176


equip__179
equip__180
equip__181
equip__182
equip__183
equip__184
equip__185
equip__186
equip__187
equip__188
equip__190
equip__191
equip__192
equip__193
equip__194
equip__195
equip__196
equip__200
equip__201
equip__202
equip__206
equip__207
equip__208
equip__209
equip__210
equip__211
equip__212
equip__214
equip__216
equip__218
equip__219
equip__220
equip__221
equip__222
equip__223
equip__224
equip__225
equip__226
equip__227
equip__229
equip__230
equip__231
equip__232
equip__233
equip__234
equip__235
equip__236
equip__237
equip__238
equip__239
equip__240
equip__241
equip__242
equip__243
equip__244
equip__245
equip__246
equip__247
equip__248
equip__249
equip__250
equip__251
equip__252
equip__253



equip__254
equip__255
equip__256
equip__257
equip__258
equip__259
equip__260
equip__262
equip__267
equip__268
equip__269
equip__270
equip__272
equip__273
equip__275
equip__279
equip__280
equip__281
equip__282
equip__283
equip__284
equip__287
equip__288
equip__289
equip__290
equip__291
equip__292
equip__293
equip__295
equip__296
equip__297
equip__298
equip__299
equip__300
equip__301
equip__302
equip__303
equip__304
equip__305
equip__307
equip__310
equip__311
equip__312
equip__313
equip__314
equip__315
equip__316
equip__317
equip__318
equip__319
equip__320
equip__321
equip__322
equip__323
equip__324
equip__325
equip__326
equip__327
equip__328
equip__329
equip__330
equip__332
equip__333
equip__334


equip__336
equip__337
equip__338
equip__339
equip__341
equip__342
equip__343
equip__344
equip__345
equip__346
equip__347
equip__348
equip__349
equip__350
equip__351
equip__352
equip__353
equip__354
equip__355
equip__356
equip__357
equip__358
equip__359
equip__360
equip__361
equip__362
equip__363
equip__364
equip__365
equip__366
equip__367
equip__368
equip__369
equip__370
equip__371
equip__372
equip__373
equip__374
equip__375
equip__376
equip__377
equip__378
equip__379
equip__380
equip__381
equip__382
equip__384
equip__385
equip__386
equip__387
equip__388
equip__389
equip__390
equip__391
equip__392
equip__393
equip__394
equip__395
equip__396
equip__400
equip__401
equip__402
equip__408
equip__410


equip__421
equip__417
equip__416
equip__415
equip__414
equip__413
equip__412
equip__411
equip__430
equip__429
equip__428
equip__427
equip__426
equip__425
equip__424
equip__422
equip__439
equip__438
equip__436
equip__435
equip__434
equip__433
equip__432
equip__431
equip__458
equip__455
equip__452
equip__451
equip__449
equip__445
equip__444
equip__443
equip__469
equip__468
equip__465
equip__463
equip__462
equip__461
equip__460
equip__459
equip__478
equip__477
equip__476
equip__475
equip__474
equip__472
equip__471
equip__470
equip__486
equip__485
equip__484
equip__483
equip__482
equip__481
equip__480
equip__479
equip__489
equip__488
equip__487