Spatial Game Analytics and -Visualization

Anders Drachen
College of Computer and Information Science
Northeastern University
Boston, Massachusetts, US
andersdrachen@gmail.com

Matthias Schubert
Institute of Informatics
Ludwig-Maximilians-University Munich
Munich, Germany
Matthias.Schubert@ifi.lmu.de

***Abstract*—The recently emerged field of game analytics and the development and adaptation of business intelligence techniques to support game design and development has given data-driven techniques a direct role in game development. Given that all digital games contain some sort of spatial operation, techniques for spatial analysis had their share in these developments. However, the methods for analyzing and visualizing spatial and spatio-temporal patterns in player behavior being used by the game industry are not as diverse as the range of techniques utilized in game research, leaving room for a continuing development. This paper presents a review of current work on spatial and spatio-temporal game analytics across industry and research, describing and defining the key terminology, outlining current techniques and their application. We summarize the current problems and challenges in the field, and present four key areas of spatial and spatio-temporal analytics: Spatial Outlier Detection, Spatial Clustering, Spatial Predictive Models, Spatial Pattern and Rule Mining. All key areas are well-established outside the context of games and hold the potential to reshape the research roadmap in game analytics.**

Keywords— Telemetry, metrics, game development, game design

I. INTRODUCTION

All digital games involve some form of spatial operations ranging from simple point-and-click vector mechanics, navigation in 2D environments like side-scroller games, to fully-fledged 3D avatar-based movement in a Massively Multiplayer Online Game (MMOG). The spatial component often forms one of the basic elements in the playing experience. Therefore, the analysis and evaluation of spatial behavior is of direct interest to game design and Game User Research (GUR): When investigating whether a game design works, the spatial dimension is one of the aspects that needs to be evaluated [5,9,20,21]. An example is the evaluation of a MMOG playfield with respect to problems that hinder or alternatively promote the intended progress of the players [5,11,20,21,29]. A component of player-derived behavioral telemetry data which is often directly tied in with the analysis of spatial data is the temporal dimension of play. All game play - occurs over time, and it is therefore common in GUR (whether based in industry or academia) to integrate this dimension, i.e. performing temporal analysis – when moving beyond simple aggregations and visualizations of telemetry data. For example, many standard Key Performance Indicators (KPIs) from social online games are based on temporal measures, e.g. Day Active Users (DAU) and User Lifetime Revenue (ULR) [33]. Similarly, the temporal aspects of player behavior are vital to progression analysis [4], trajectory analysis [18], bot detection analysis [2,21,25,27], time spent analysis, and many other forms of analysis based on player telemetry. In this paper, the term spatial analysis will denote both spatial and spatio-temporal visualization and analysis techniques. Spatial data mining methods explicitly consider the spatial component of the data like movement, often in addition to non-spatial components like character attributes [11]. Thus, it is possible to explicitly distinguish information about which objects are located close on the map and which objects are similar based on non-spatial attributes. Finally, spatio-temporal data mining allows considering the change of all of these types of information over time.

Spatial analytics is generally carried out in a knowledge discovery process [3,7,30,33] including data mining and machine learning techniques. The goal is to extract patterns which are statistically correct and useful. A pattern in this context ranges from simple rules (e.g. "players will follow the road") to complex mathematical functions. The knowledge that can be extracted from spatial data is often more complex than patterns being derived from non-spatial data. Thus, specialized methods may be required to consider spatial relations. Basic visualizations such as heatmaps are common, e.g. for major titles such as Halo 3 and Team Fortress 2, and visualizations of event-based player telemetry on top of maps of game environments has been showcased in several presentations and articles [e.g. 5, 9, 29]. More complex spatial analytics has been applied within game-oriented AI research, e.g. for developing lifelike bot behavior in First Person Shooters (FPS) or bot detection in MMOGs [e.g. 19, 27]. The player community has also been active in generating spatial visualizations of data, e.g. tools for resource gathering in World of Warcraft (such as: www.wowhead.com).

In this paper, we review spatial and spatio-temporal game analytics, including definitions of key terms and techniques. The current directions in spatial and spatio-temporal data mining for digital games are outlined, including a series of industry- and research-based example cases. Potential future directions for spatial and spatio-temporal game analytics are outlined and possible application areas discussed. The article does not aim to be all-inclusive but to provide a fundamental review for and of spatial game analytics (for an in-depth overview of all directions in spatio-temporal analytics, see e.g. [31]). For example, the challenges of collecting and storing telemetry data are left out. We will focus on player-derived telemetry recorded at clients or servers, i.e. systems telemetry which is useful for monitoring game systems and network balancing [32], is not discussed here.

II. Terms and Definitions

The terms “spatial analytics” or “spatio-temporal analytics” in the context of computer games are usually describing the evaluation or analysis of the spatial component of player behavior, obtained via telemetry sources [5]. For example, the X, Y, and Z coordinates of the location of a player, as well as the time, whenever that player dies, fires a weapon, accepts a quest, punches an opponent, etc. –when an event of any type occurs. In some types of games, it may not make sense to use specific coordinates, but rather information about which zone or area the player is in when a given event occurs. For mobile games, the coordinates of the position of the player in the real world can also be of interest. Irrespective, these are all examples of spatial information that come attached to specific player behaviors. In practice, spatial analytics often involves a temporal dimension as well, and the two terms “spatial” and “spatio-temporal” are often used to denote the same analytical approach. Another important aspect of the terminology is the difference between analysis and visualization. Spatial analysis involves performing calculations on data with a spatial component, for example multiplication, addition or subtraction. Visualization does not involve calculations, but can involve descriptive or aggregative manipulation.

Four types of information can be logged whenever a player does something – or is exposed to something – in a game: Who is it happening to? What is happening? Where is it happening? At what time is it happening? More formally, a data object describing a player avatar in a virtual environment can be described by four types of information: 1) Physical attributes of the avatar (or avatars – there are various ways in which the player/-s can be visually expressed in games, if at all) and the abilities that it encompasses, e.g. for a typical MMORPG class, level, health, strength, speed and similar. 2) The involved event or action, e.g. the used abilities, its target etc. 3) The spatial position, movement speed and current direction of the avatar. In case of games where no player avatar exists, only the second category comes into effect (e.g. position and direction of swipes on a touch screen). 4) All of these types of information can change as a function of time, adding a temporal component [5,20,27]. We can refer to these components of player-derived telemetry data as Avatar, Event, Space and Time information. A potential fifth component is Social, i.e. “to whom”. However, the social aspects of telemetry analysis is less well researched in the spatial domain and therefore not discussed further here. Spatial information generally comes in three types: points (coordinates), lines and areas (polygons) [3]. Point-based data is typically defined using coordinate sets (X,Y and possibly Z). Lines are spanned by multiple points and can integrate information about direction. Area data has a spatial extension, for example a building or area with a particular type of vegetation.

Spatial measures as associated with attributes: Space is always measured on conjunction with an event (and usually also a time). This also means that when dealing with games such as an FPS, player behavior can be mapped as: 1) A spatio-temporal trajectory; 2) the sets of actions performed by or performed on the player, each associated with a time stamp and some spatial

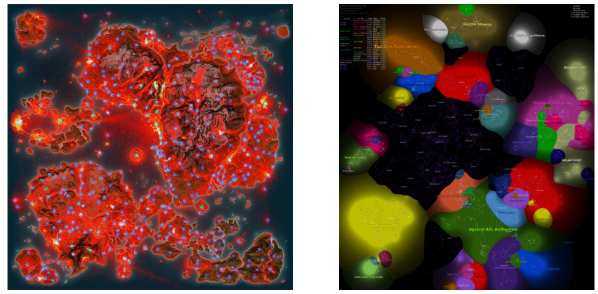

Figure 1: Spatial visualizations. left: A faction control map from *EVE Online*. Source: [26]; right: A heatmap of extraction events in *Just Cause 2*. More than 22.3 million events form the basis for the heatmap. Color ramp is scaled to the cell with the largest value is always white. Source: [1]

information[2]. In addition to player-derived telemetry, system-derived telemetry can be logged from game clients/servers. We refer to these two different sources of data as player telemetry and system telemetry. The majority of spatial analytics performed in games is based on player telemetry. However, playing experience is shaped by the interaction between player and game and system telemetry is therefore also relevant to evaluate, e.g. the behavior of AI-controlled agents, trajectories of mobile entities and pathing routines (dynamic object tracking), bug tracking, etc. [5,19,27].

III. Tools for Spatial Analysis

There is a wide variety of tools available for performing spatial game analytics. In recent years, various start-up companie have begun offering technologies for tracking, logging and storing game telemetry data (e.g. Honeytracks, Tableau, Game Analytics, Playnomics). The tools offered by these companies remain in their infancy in terms of breadth, compared to the commercially available packages for data mining, web analytics and spatial analytics. In order to perform most kinds of spatial analytics on game telemetry data, users need to look outside game-specific solutions or alternatively develop the tools in-house, e.g. the “TRUE” system developed by Microsoft Game User Research Labs [11], “Skynet” developed at Bioware [29], “Data Cracker” [17], “G-player” and “GA Tool” at Game Analytics or “Lithium” developed by Hoobler et al. [9]. Developed specifically for spatial analytics, Geographic Information Systems (GIS) form the major platform for managing spatial information [3]. A GIS is conceptually similar to the data management systems used in games which control objects and entities with specific attributes and behaviors inside the game environment. The difference is that a GIS is specifically designed for the management of spatial and spatio-temporal data. In a GIS, map features are linked with attribute information. For example, a level map is linked with locations of quest providers, different types of environments, player trajectories, or other spatially distributed features. In a GIS, player behavior metrics derived from telemetry data (also referred to as “gameplay metrics” [e.g. 5, 33]) can be visualized on top of or inside the game environment. For example, player trajectories can be imported, plotted on top of a level map, and then analyzed in terms of speed of progression through the game´s environment (and whether this matches design expectations).

IV. CURRENT DIRECTIONS

A. *Univariate/bivariate map analytics and visualizatons*

Spatial game metrics can be visualized via synthesis without any accompanying analysis phase. Univariate (one variable) and bivariate (two variables) visualizations are among the most commonly described techniques in the game development and game research literature [e.g. 5, 11, 20, 29] due to their easy generation and intuitive interpretation. Visualizations of this type have also been generated by players, e.g. with the goal of providing guidance or strategic advice. Examples include resource maps for MMOGs and open-world games such as Fallout 3 and Oblivion. Developers similarly generate visualizations not only for internal use, but also for the player community, e.g. in the form of global heatmaps for Halo 3 (http://www.bungie.net/Online/HeatMaps.aspx). Furthermore, similar visualizations are used internally to evaluate designs. For example, to check whether any mobile entity wanders outside the area it is supposed to stay within or where players pick up items. In general, plotting where something happens on a map provides a context that a non-spatial analysis will not provide (Figure 2). Spatial visualizations provide a means of evaluating gameplay telemetry (gameplay metrics), in the context that the games are actually played in. This adds an explanatory dimension that is not available in non-spatial analytics (e.g. causes of death as a function of location; Figure 2). Almost all game metric visualizations to date have been generated based on point features (an exception being e.g. [5], (Fig. 2), with line and area features remaining less used. However, in principle these types of spatial features are equally applicable.

Generating heatmaps can be performed using aggregate techniques or via several other types of algorithms. The technique originated in 2D displays of the values in a data matrix, but has been adapted to a variety of contexts like heatmapping the eyegaze on websites or mapping environmental factors. A typical way to generate a heatmap is to divide the game area under scrutiny into a grid of cells (bins), and sum up the number of events that occur within the area covered by each cell. Adding a color ramp (e.g. green to red) allows easy interpretation of the data. Other approaches to generate heatmaps employ kernel density functions or alternatively interpolation in order to attempt simple prediction in areas that there are no events registered for [3, 5]. In games, heatmaps are most often used to aggregate and visualize death events, forming maps of the lethality [1, 5, 9]. Analyzing the spatial distribution of different causes of death adds a second layer usefulness to heatmaps, as it is thus possible to evaluate the impact of different enemies or weapons in different regions of an area. Most available game heatmaps have been based on 3D-games using point data (X,Y); however, applications for generating 3D heatmaps (X,Y,Z) exist and allow for better interpretation of the effect of level design. Thus, 3D heatmaps help to alleviate the errors that can occur when data from multiple Z-axis levels (e.g. a building with two floors) are layered on top of each other.

Interpreting heatmaps requires no special training, and allows for ready identification of e.g. bottlenecks [5]. It is

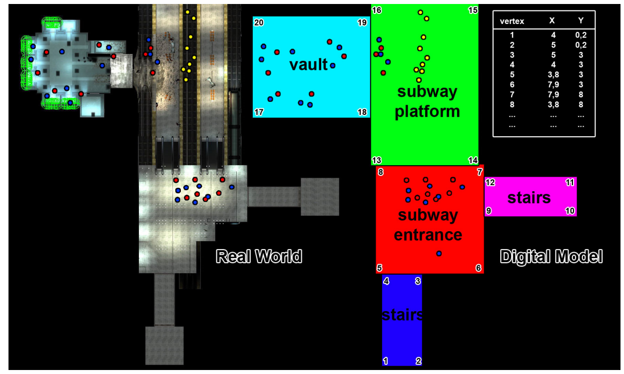

Figure 2: An example of the construction of a polygon-based model of a "real world" (i.e. a 3D game environment) on top of a game level map (digital model). Locations of player death events forms a third layer (point events). Combining the polygon-based model and the death events layer provides the ability to perform calculations across these two layers, for example to calculate how many death events that fall into each of the major areas of the game level. Source: Seif El-Nasr et al. [33]. Illustrations by Dr. Alessandro Canossa.

important to ensure that heatmaps are highly granular, enabling detailed analysis of a playfield design.

Heatmaps are also useful to evaluate whether any area of a map is not being used/experienced, although trajectory data are more precise for this purpose [20]. Importantly, heatmaps are not limited to the mapping of death events, nor to the mapping of point-based data. The simple aggregation process means that e.g. player trajectory data (line data) can also be used as the basis for heatmap visualizations, showing in these cases the density of chosen movement paths. Similarly, area data (polygon data) can be aggregated. Finally, combining heatmap visualizations with the temporal dimension of telemetry data adds a dynamic quality and allows for a better understanding of game flow. Essentially, heatmaps combined with temporal information allow designers to follow how the mapped events occur over time. Thus, it is not difficult to think of situations where it is interesting to know e.g. what players do within the first 30 seconds on a map as opposed to a simple aggregate of the sum of all of their activities [5, 26].

B. *Multivariate analytics and visualizatons*

Visualizing and analyzing multiple spatial features (variables) enables us to evaluate how these variables interact [3,5,9]. For example, locations of player death combined with trajectory data provide a highly detailed analysis of the dynamics of a playfield. Mapping multiple variables is in the context of a GIS approached via a process known as overlapping. This mapping principle has also been applied to game analytics [e.g. 5,10]. Overlapping is a spatial operation in which two or more maps or layers are registered to a common coordinate system being superimposed on top of each other [3]. The overlay operation thus creates composite maps by combining datasets. The purpose is to visualize and analyze the relationship between features occupying the same space, e.g. a map of a game level, player trajectories logged for that level, and death events. The overlay function can be based on simple operations such as synthesis (trajectories + where players die), or analysis (e.g. multiplying, subtracting, find averages or co-occurrences – for example subtracting two heatmaps from each

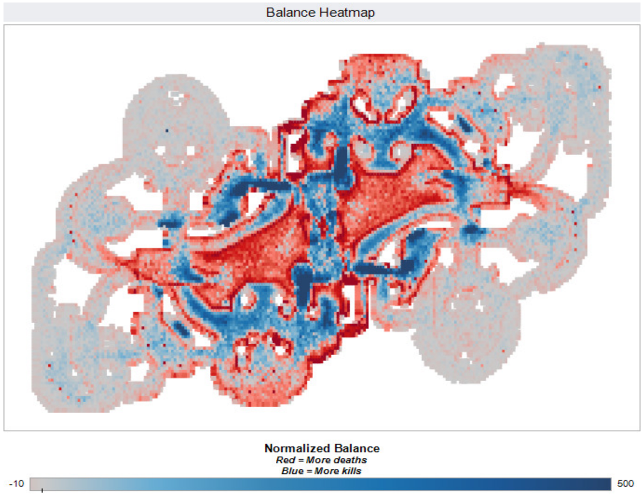

Figure 3: Example of overlay analysis: A balance heatmap from the "Molten" map of the game *Transformers: War for Cybertron*, generated using an overlay function, where a death heatmap is subtracted from a kill heatmap. Areas with negative values (red) indicate dangerous areas; areas with positive values (blue) indicate areas that are safer. A conclusion is that wall areas appear to be dangerous, possibly due to restricted movement. Source: Houghton [10].

other to evaluate the difference). Spatial data can be displayed as raster or vector models. Raster models are grids – each grid cell contains information about one or more variables; vector models are comprised of point and lines. Raster and vector models differ significantly in the way overlay operations are performed, and generally overlay operations are performed most effectively in raster-based models [3]. Drachen & Canossa [2011] describes an example of overlay analysis of areas where players die of different causes; to locate areas of a level in *Tomb Raider: Underworld*. In this game, multiple threats to the players are present. Correlating threats with death events it can shown that areas with multiple threats are more lethal than areas with a single type of threat. An important point to be aware of when performing overlay operations in 3D games and projecting these to a 2D plane, is that most of these games have variations in the Z-plane, leading to different opportunities for players to position themselves (e.g. elevated positions). Levels may even have multiple planes on top of each other (e.g. floors in a building) which adds bias, and should ideally be kept separate. Similarly, some visualized patterns may not be clear unless the 3D nature of the playfield is considered. The problem can be solved by using 3D visualizations of the data, something that is still rare in games (but see e.g. [1]), but potentially a powerful tool for designers for evaluating player behavior in the dimensional space available to them. Accounting for the temporal nature of behavioral data adds another layer of information.

C. *Trajectory analysis*

A trajectory is a description of the movement of an object in space over a specific period of time, for example the navigation of a player through a game level. Along the way, various events occur – fights, item pickups, item uses etc. [37] Trajectory analysis is thus useful for getting close to the actual gaming

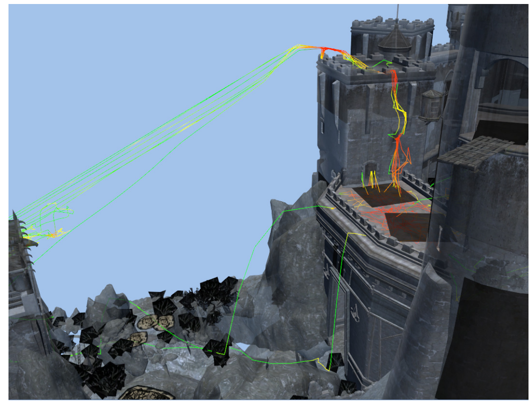

Figure 5: Trajectory visualization developed by the user research team at Ubisoft Montreal using DNA Viewer, an in-house telemetry system. Eight player trajectories are shown, based on a playtest from *Assassins Creed 2*. Six players are shown parachuting from the castle tower (right side of image), but two choose an alternative strategy along the castle walls, which was unintended by the design as it bypasses the tutorial and results in a less engaging experience. The visualization was generated per the request of the user research team, who noted the difference in player behavior. Source: Dankoff [39].

experience. Analyzing trajectories is currently used to locate illegal bot programs in online games, examine group behavior, study player tactics, asset use etc. Trajectories are in practice approximations. There are various ways to record spatial movement, for example as a series of point positions (waypoints) ordered according to time played, and lines drawn to connect them. Depending on the frequency of waypoint logging, it can be a good idea to measure the relative speed of the player together with the position, in order to estimate e.g. which sections a player is running vs. walking through. Trajectories are among the more bandwidth-heavy player behavior features to track and a degree of filtering can be a good idea to avoid trouble with transferring and storing telemetry data. Another way to describe trajectories is as changes in relative movement, i.e. movement stored as a sequence of commands such as "go straight", "turn right", "stop" etc. This form of pattern logging is currently used in research focusing on distinguishing between bots and players. Trajectory analysis has established types of patterns which are indicative of specific trajectory behaviors like tracks, flocks and leadership patterns [27]. A track describes a player having a constant movement in a given spatial range, e.g. a set of players moving on a straight line within a certain area form a track. A flock describes a set of players travelling together in a way that all members are within a certain diameter or spatial range at each point of time. In a leadership pattern, players might join the movement of a leading player at some point in time [28]. One form of trajectory analysis in games is clustering, i.e. classification of data based on the degree of similarity and dissimilarity. Clustering can be based on plain trajectories or trajectories joint with event data, like ability usage, item pickups or player interactions. Combining trajectories with event data is useful because paths alone do not necessarily tell us why players navigated in a particular way or whether they moved together. Usually cooperative behavior is limited to a specific period of time, which can make spatio-temporal data analysis interesting. Clustering trajectories requires a measure of similarity between trajectories [28]. Applying data mining techniques for trajectory data has been already employed for various use cases in game analysis [18,19,27]. In Miller & Crowcroft [18] the authors examine character movements in World of Warcraft. In particular, the analysis is concerned with character movement in the Arathi Basin player-vs-player battle ground (Figure 6). The authors examine three types of information: 1) Waypoint-based movement models. In this type of analysis trajectories can be shortened to the part connecting two waypoints. Analysis indicated that one group of players, referred to as patrollers, travel between a small set of waypoints and use efficient routes on their travel, trying to capture or hold nearby bases; 2) Spatial hot spots. A hot spot is an area of the map where characters spend most of their time. To derive hotspots the authors partition the map into a grid and record the timethe characters spend in each grid cell. The cells with the highest counts are selected as hotspots. The analysis of hot spots revealed a second type of player: guards which stays close to a base for the complete game to defend it. 3) Flocking and grouping of avatars. The authors also recorded characters moving together within a 30 yards diameter for several seconds. This analysis did not reveal typical group movement. This can be explained by dying players which respawn at the nearest grave yard and rejoin different groups of other players.

A well-examined topic in academia-based game analysis is detecting bot programs controlling an avatar in a MMOG instead of a human player. Since movement is in many games the most common action of avatars and units, it is possible to distinguish bot controlled from human players by analyzing their movement in the virtual environment. In Pao et al. [19], the authors propose to detect bots by analyzing trajectory information in the FPS Quake II. The authors propose two representations to describe the characteristics of player movement: 1) Distribution of the step size; 2) Measuring the change in step size and direction relative to the previous point of time. The results achieved more than 96% of prediction accuracy for detecting three types of common Quake II bots based on trajectories corresponding to 100 seconds.

Trajectory analysis might include the type of activity a moving object is currently undertaking. In geo-information systems a lot of the necessity for deriving this type of information is based on the incompleteness of the available data. A classical task in this area is transportation mode detection, e.g. to predict whether a cell phone owner travels by bus, bike or car. In a game, comparable information is usually known. However, the employed methods can be extended to mine complex behavioral patterns. In gaming applications, activity recognition allows for combining spatio-temporal movement data with activity patterns. Examples of the application of behavioral patterns in a computer game include mining build orders in the real time strategy games such as Starcraft [30]. A build order describes the order of buildings and units that has to be followed to pursuit a certain strategy. For example, a player wanting to start an early surprise attack on his opponent will select a strategy favoring the training of attack units as early as possible instead of collecting a lot of resources first. Analyzing build orders yields valuable information about the game balance: Since the opening moves in most games tend

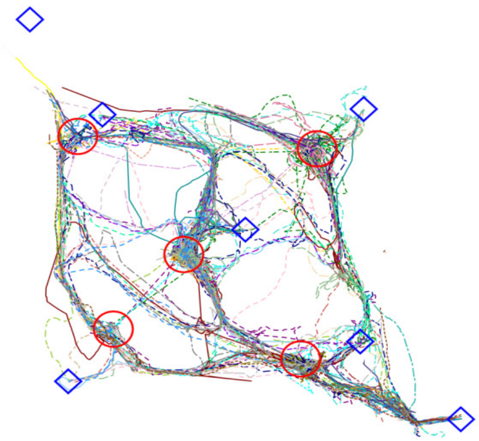

Figure 6: Left: 980 player trajectories from the Arathi Basin battleground of *World of Warcraft.* Red circles indicate strategic points; Source: Miller and Crowcroft [18].

to follow a rather static pattern, most players can follow them easily. Thus, finding a starting strategy without a proper counter indicates a flaw in game balance [30].

D. Behavioral analytics

The fourth category represents a mixed bag of methods focused on the overall goal of delivering behavioral analytics, and has so far primarily been applied in game research. However, commercial games like StarCraft, and Quake III were used as test bed. Notably the area of bot behavior (e.g. goal planning, movement path planning) has traditionally been a common goal between academic research and industrial development, indicating a potential for future collaboration. While approaches/applications for behavioral analytics are quite heterogeneous, there are two things all have in common:

1) Behavior is considered as a successive sequence of actions. Thus, behavior is considered a set of several single actions that are done to achieve a certain goal.

2) The behavioral context is considered. Different situations will lead to different types of behavior. An example is the model for modeling human-like player behavior described in [27]. By clustering the player status and connecting it to the subsequent behavior, the model distinguishes first the context and then, describes the movement. In particular, the behavior of a player running out of ammo will be to collect new ammo, while the behavior of a player recognizing that his opponent has no ammo left will be to hunt him down while he is more or less defenseless.

Methodically, the tool box of describing behavioral patterns contains e.g. methods for sequential data analysis [8] such as sequential pattern mining and Markov Chains [33]. Both methods are applicable to cases where all the actions describing the behavior can be directly observed. If these steps are not directly accessible, methods like hidden Markov models, latent variable analysis or conditional random fields might be applied [28]. For Markov chains and sequential patterns, it is assumed to make observations being connected to the actual behavior. Though, it might appear that most applications in game analytics are based on clearly observable actions, there are as well use cases requiring the more complicated observation based models. One reason for the use of models distinguishing between the observed actions and the intended ones is fault tolerance. A player showing a complex action pattern might tend to make mistakes. Thus, the observed action might not be the intended one. Furthermore, observation based models are often more tolerant with variations of less important parts of a behavior. Finally, in some games some of the player actions cannot be observed because they are not done in the game. An example, are poker bots calculating win probabilities offline. Furthermore, team communication in computer gaming is often done by external voice chats. A final technical reason might also be that storing all actions in an online game for later analysis might cost too many resources. Thus, the game analytics expert has to suffice with information being currently collected, even though it might be possible to extract and store more detailed information.

There are currently several papers describing player behavior as sequential patterns, e.g. Breining et al. [2] applies sequential pattern mining to modeling player behavior. The approach describes player behavior as a sequence of the performed player actions. Additionally, two extensions are introduced: 1) Considering a time step of each action allowing for distinguishing how fast a certain action sequence was performed; 2) Adding a spatial position to each action. Thus, it is possible to model trajectories and see whether there is a certain unknown correlation between actions and locations on the map. The experiments were performed on Quake III data and showed successful results for player identification and finding local behavior patterns.

A system modeling player behavior as sequential patterns is SABAF (Sequence Alignment Analysis of Player Behavior) [25]. The task of SABAF is to predict whether players will be inactive for at least 30 days for the MMORPG Everquest II. To describe the player activity the method describes players by a sequence of 25 actions, e.g. finished quest, killed a monster or was killed. Based on this description players are compared with local and global sequence alignment. To make a prediction SABAF determines the most similar players in sample database containing active and inactive players. Based on the most similar cases in the sample database the system decides whether it is likely that the player will return to play the game or just quits. The results indicate that the order in which events occurred seems to be important to the behavior of the player.

The methods being discussed so far describe the behavior of game entities in an isolated view, i.e. the behavior is categorized and judged only with respect to the intentions and the actual actions of the player. However, most relevant player actions are strongly dependent on the behavior of other players. To judge whether a strategy works out or not, we have to consider what the other players are doing. For example, choosing a proper opening in a real-time strategy game is strongly dependent on what opponents and team mates are doing. Mining collaborative and antagonistic behavior is a very complex challenge without a lot of general solutions so far. The general problem is to decide which actions should be considered as related and which are not. By treating everything as related to everything else, we usually run into an over-fitting view of the world displaying nothing but unique events instead of universal behavioral patterns. On the other hand, the key component of team play might only be described by complex spatial relationships [12].

V. FUTURE FLAGSHIP DIRECTIONS

Four areas with the potential to reshape the roadmap in spatial game analytics. All are established areas of R&D outside digital game. The list is synthesized from current knowledge and should not be considered inclusive, but a starting point for future research.

Spatial Prediction: Prediction via ML techniques has become a much-discussed and much adopted technique in game development [33, 34], perhaps notably because of the shift in business models from retail to online and F2P. When a game´s revenue is dependent on micro-sales of virtual currency and – items, being able to predict which players that will buy, why and when, becomes useful for driving revenue [34]. As yet, predictive analytics using spatial dimensions have not been published by the game industry. Spatial predictive models are outside games used to find spots or areas on a map where certain events might occur. Example use cases are predicting crime spots, spots for natural disasters like flooding or predicting the breeding spots of birds. In general, predictive models learn the characteristics of locations where certain events already occurred and employ this knowledge to predict the likelihood that the same event happens at another location. Common methods for spatial prediction include Markov random fields [14, 22] and spatial autoregressive models [22]. Both methods are based on statistics and include the characteristics of a certain area in combination with its surrounding. **Potential:** In game development, spatial predictive models can possibly be applied to predict server load across different in-game areas [e.g. 32], as well as estimate the response in the spatio-temporal behavior of players to design changes, notably in MMOG contexts. An example of spatio-temporal behavior changing due to a design change was shown by Canossa et al. [37]. Furthermore, spatial predictive models add a layer of information to non-spatial predictive models [16], providing contextual information which would otherwise not be available. Thus, they assist in determining the causal drivers of behavior, e.g. players quitting a game.

Spatial Outlier Detection: In general, an outlier is defined by a data object which does not fit to the data distribution holding for the rest of the data set. Spatial outlier detection specializes this concept by comparing an object to its spatial neighbors. For example, a house having a moderate price in the middle of a high class residential area yields an interesting spatial outlier. In this example, the spatial component is essential to the interestingness of the object because there might be other estates with a similar price. **Potential:** In game development, spatial outlier detection could be used to find exploitation spots. Notably, there might be spots on a map that allow for attacking computer opponents without having them defending themselves. Example techniques for spatial outlier detection are described in Lu et al. [15]. Exploitation – and cheating - is notably a problem in persistent-world games, but finding weaknesses in level design which provides unbalanced “sweet spots” is a common challenge in game user research [33].

Spatial Clustering: Spatial clustering describes the grouping of spatial objects that are similar with respect to their general characteristics and additionally, are positioned close to each other on the map. In general, objects from different clusters should be as dissimilar as possible. A general application of spatial clustering is to find out whether spatial objects are uniformly distributed across an environment or if there are hot spots where e.g. certain events happen more often. An overview to clustering algorithms suitable for spatial data can be found in [7]. **Potential:** Clustering has been applied to find patterns in player behavior. During production it can be used to evaluate whether a design offers the planned-for affordances and after the launch, it is used to evaluate how the game is played. For example, Drachen et al. [2009] and Drachen et al. [2012] applied different unsupervised techniques to cluster players in major commercial games such as Tomb Raider: Underworld and Battlefield 2. Combining methods for player categorization with visualization of trajectories inside game environments provides an accessible form of feedback to game designers. However, much work is needed in the area, notably in terms of dynamic and interactive visualizations [17,33,36].

Spatial Co-Location Patterns and Spatial Rule Mining: These techniques focus on deriving rules about spatial co-location and other relationships between spatial locations. In the case of spatial co-location rules [23], we are interested in pairs of certain objects occurring nearby each other. A general application in Biology is the co-existence of certain animal species. Methods for spatial trend detection are described in [6]. In general, an association rule is considered a spatial association rule if the predicates being used for formulating the rule describe spatial relationships or characteristics. An algorithm for deriving spatial association rules can be found in [13]. **Potential:** In games, spatial co-occurrence rules can be used to derive higher order tactical elements like group compositions in a strategy game. Another method resulting in spatial rules is spatial trend detection, concerned with the change of features when moving into a certain direction, e.g. increasing housing prices between suburbs or probabilities of specific events happening in a game environment. Spatial association rule mining conversely describes a probabilistic implication. An example, describing an analysis of the time players spent in specific game levels of Tomb Raider: Underworld, noted that players spending too much time in a specific early area of the game will have a higher chance of quitting the game early [16].

VI. OPEN PROBLEMS

The use of player behavior telemetry in academic games research and industrial development has increased within the past few years. The current push in this direction is championed by biggest game publishers on the industry side [33], notably within the fields of game user research and game analytics; and in academia by fields such as user experience [36] and game AI [35]. While this interest is pushing innovation, game analytics is a young field, and for spatial analytics specifically, a number of fundamental challenges need to be addressed, including:

1) The confidential nature of player-derived telemetry data and the results obtained from analysis thereof has proven an effective barrier towards knowledge sharing across the game sector, and to building academic-industry partnerships [33]. This may change as big publishers such as EA, Ubisoft, Square Enix and others have started collaborations with researchers on developing new techniques for analysis and visualization of player behavior telemetry [e.g. 33,35].

2) Lack of awareness of how to apply spatial game analytics to inform game development in the game industry and outside the specific research networks that utilize behavioral telemetry already, e.g. game user research and game AI. Essentially, if a game development company is not aware of spatial analytics and its potential applications, it will not investigate the topic. Game analytics in general is a topic gaining increasing attention at industry events which may help drive awareness in the coming years [29, 33], but there is currently a lack of knowledge about methods, techniques and their application to improve games.

3) The distributed nature of the relevant research utilizing behavioral telemetry, e.g. networking [32], game AI [35], game user research [33, 36], data mining [27], and interactive storytelling. The distributed nature means that even published knowledge does not always cross the borders between the involved fields of research. Associated with this issue is the fact that spatio-temporal analysis is used across a wide range of industries and research fields, e.g. mining, urban planning, satellite image analysis, environmental monitoring, geomarketing etc. [e.g. 3,6,13,23,24]. In these areas, spatial analytics has a decades-old history, and given the similarities between at least some virtual environments and the real world, there would appear to be the potential for finding techniques in these fields and adapting them to games, as has been done e.g. with trajectory analysis [e.g.18,19,27].

4) Lack of software tools which enables non-experts to perform spatio-temporal analyses and to visualize the results in a manner targeted at stakeholders in game development. While developing heatmaps is routinely done in game development [33], more advanced methods such as trajectory analysis are only reported applied to behavioral telemetry by large game publishers [e.g. 29, 36, 39], or by academics [e.g. 9] – sometimes collaborating with the industry [e.g. 17]. Adopting GIS platforms like QGIS, ArcGIS and MapInfo, while granting access to advanced spatial analytics, requires training, which can be difficult to set aside resources to for a small or medium size game development company.

5) Massive data size. Depending on the specific usage context, data sizes for spatial and spatio-temporal analytics can quickly reach TBs in size. For example trajectory logging takes substantial bandwidth and storage space due to the high frequency of logging player locations [5, 29]. This in turn increases the resource demand on storage space and processing power, which can be a barrier especially to small and medium-size game developers; as well as academics. Adopting careful sampling strategies forms a venue for partly alleviating this problem [33].

VII. DISCUSSION AND CONCLUSION

In this article a brief review of the SOTA of spatial game analytics has been provided, including definitions of key terms and techniques. Current directions in spatial game analytics work across industry and academia have been described and categorized, and numerous case examples presented. There is however research available that ties in with these topics which there has not been space to cover here, e.g. advances in adaptive gaming [35]. Spatial game analytics can be advantageous in comparison with traditional non-spatial analysis because it deals with the perspective that games are experienced through. It also allows analyses which are not possible in the non-spatial domain such as trajectory analysis, useful for e.g. determining asset use [18, 33]. There are a substantial number of challenges and open questions, including a lack of distributed knowledge about which solutions to use in which cases, and the high barrier of entry for non-experts. The current innovation in game analytics is probably going to change this in the coming years as more and more tools become available. Finally, methods for describing/mining collaborative and antagonistic behaviors are still in the development. Spatial game analytics – and analytics in general – have a potential for transforming game development in much the same way it has done in sectors such as marketing, mobile applications and environmental modeling, however, games are unique beasts and it remains to be seen how big an influence spatial analytics will have.

REFERENCES

[1] Blackhurst, J. 2011. Heatmaps, point cloud and big data in processing. URL: http://jimblackhurst.com/wp/2011/05/17/ heatmaps-point-clouds-and-big-data-in-processing/.

[2] Breining, S.; Kriegel, H.-P.; Züfle, A. & Schubert, M. 2011 Action Sequence Mining. In ECML/PKDD Workshop on Machine Learning and Data Mining in Games.

[3] Demers, M. N. 2008. Fundamentals of Geographical Information Systems. Wiley & Sons.

[4] A. Drachen, A. Canossa, and G. N. Yannakakis, 2009. Player Modeling using Self-Organization in Tomb Raider: Underworld. In Proc. of the IEEE Symposium on Computational Intelligence and Games (Milan, Italy). 1–8.

[5] Drachen, A.; Canossa, A. 2011. Evaluating Motion: Spatial User Behavior in Virtuel Environments. International Journal of Arts and Technology (4,3), 294-314.

[6] Ester, M.; Frommelt, A.; Kriegel, H.-P. & Sander, J. 1998. Algorithms for Characterization and Trend Detection in Spatial Databases. In Proc. of the 4th ACM Int´l Conf. on Knowledge Discovery and Data Mining (New York City, NY).

[7] Han, J.; Kamber, M. & Tung, A. 2001. Spatial Clustering Methods in Data Mining: A Survey. In Geographic Data Mining and Knowledge Discovery. Taylor and Francis.

[8] Han J., Kamber M. 2006. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers.

[9] Hoobler, N., Humphreys, G. and Agrawala, M. 2004. Visualizing competitive behaviors in multiuser virtual environments. In Proc. IEEE Visualization Conference.

[10] Houghton, S. 2011. Balance and Flow Maps. #AltDevBlogADay . URL: altdevblogaday.com/2011/06/01/balance-and-flow-maps-2/

[11] Kim, J.H., Gunn, D.V., Suhuh, E., Phillips, B.C., Pagulayan, R.J. & Wixon, D. 2008. Tracking real-time user experience (TRUE): a comprehensive instrumentation solution for complex systems. In Proc. CHI (Florence, Italy).

[12] Kim, H.-C.; Kwon, O. & Li, K.J. 2011. Spatial and Spatiotemporal Analysis of Soccer. In Proc. of 119th ACM SIGSPATIAL Int´l Conf. on Advances in GIS.

[13] Koperski, K.; & Han, J. 1995. Discovery of Spatial Association Rules in Geographic Information Databases. In Proc. Fourth International Symposium on Large Spatial Databases (Maine, USA) 47-66.

[14] Li, S. Z. 1995. Markov Random Field Modeling in Computer Vision. Springer Publishers.

[15] Lu, C.T.; Chen, D. & Kou, Y. 2003. Algorithms for spatial outlier detection. In Proc. of 3rd IEEE Inl. Conference on Data Mining.

[16] Mahlman, T.; Drachen, A.; Togelius, J., Canossa, A. & Yannakakis, G. 2010. Predicting player behavior in Tomb Raider: Underworld. Proceedings of CIG. 178-185.

[17] Medler, B.; John, M. & Lane, J. 2011. Data cracker: developing a visual game analytic tool. In Proc. Human Factors in Computing Systems (Vancouver, Canada). 2365–2374.

[18] Miller, J., L. & Crowcroft, J. 2009. Avatar Movement in World of Warcraft Battlegrounds. Proceedings of Netgames. IEEE Publishers, ISBN: 978-1-4244-5605-5.

[19] Pao, H.-K.; Chen, K.-T. & Chang, H.-C. 2010. Game Bot Detection via Avatar Trajectory Analysis. In Proc. Computational Intelligence and AI in Games, IEEE Transactions on (2,3), 162-175.

[20] Ramero, R. 2008. Succesful Instrumentation. Presentation at the Game Developers Conference.

[21] Schubert, M.; Kriegel, H.-P. & Züfle, A. 2011. Managing and Mining Multiplayer Online Games. In Proc. SSTD. 441-444.

[22] Shekhar, S.; Schrater. P.R.; Vatsavai, R.R; Weili,, W. & Chawla, S. 2002. Spatial contextual classification and prediction models for mining geospatial data. IEEE Transaction on Multimedia (4,2).

[23] Shekhar, S. & Huang, Y. 2001. Co-location Rules Mining: A Summary of Results. In Proc. of the 7th International Symp. on Spatial and Temporal Databases.

[24] Shekhar, S.; Lu, C. & Zhang, P. 2003. A Unified Approach to Detecting Spatial Outliers. GeoInformatica (7,2).

[25] Shim, K. J. & Srivastava, J. 2009. Sequence Alignment Based Analysis of Player Behavior in Massively Multiplayer Online Role-Playing Games. In Proceedings of the IEEE ICDM-10 Workshop on Domain-Driven Data Mining.

[26] The EVE Online Wiki. URL: http://wiki.eveonline.com /wikiEN/images/d/d0/Territorial_maps.png.

[27] Thurau, C.; Bauckhauge, C. & Sagerer, G. 2003. Combining SOM and multilayer perceptrons to learn bot-behavior for a commercial game. In Proc. GAME-ON Conference. 119–123.

[28] Y. Zheng and X. Zhou. 2011. Computing with Spatial Trajectories. Springer.

[29] Zoeller, G. 2010. Game Development Telemetry. Presentation at the Game Developers Conference 2011.

[30] Weber. B. and Mateas, M. 2009. A Data Mining Approach to Strategy Prediction. In IEEE Symposium on Computational Intelligence in Games (Milan, Italy). 140–147.

[31] N. Andrienko & G. Andrienko (2005). Exploratory Analysis of Spatial and Temporal Data. Springer.

[32] Feng, W.-c., Brandt, D. and Saha, D. 2007. A Long-Term Study of a Popular MMORPG. In Proc. Netgames (Melbourne, Australia).

[33] Seif El-Nasr, M., Drachen, A. and Canossa, A. 2013. Game Analytics – Maximizing the Value of Player Data. Springer.

[34] Fields, T. and Cotton, B. 2011. Social Game Design: Monetization Methods and Mechanics. Morgan Kauffman Publishers.

[35] Yannakakis. G. N. Game AI Revisited. In Proceedings of ACM Computing Frontiers Conference, 2012.

[36] B. Medler. Play with data – an exploration of play analytics and its effect on player experiences. PhD-Thesis, School of Literature, Communication and Culture, Georgia Institute of Technology.

[37] Canossa, A., Drachen, A. and Sørensen, J. R. M. 2011: Arrgghh!!! – Blending Quantitative and Qualitative Methods to Detect Player Frustration. In Proceedings of the 2011 Foundations of Digital Games Conference (Bordeaux, France).

[38] Drachen, A., Sifa, R., Bauckhage, C. and Thurau, C. 2012. Guns, Swords and Data. In Proc. of IEEE CIG (Granada, Spain), 163-170.

[39] Dankoff, J. Game Telemetry with Playtest DNA on Assassin´s Creed. The Engine Room, September 12, 2011. URL: http://engineroom.ubi.com/game-telemetry-with-playtest-dna-on-assassins-creed.