Developing New Approaches for the Analysis of Movement Data

A sport-oriented application

Pengdong Zhang

The research reported in this dissertation was conducted at the CartoGIS research unit, Department of Geography, Faculty of Sciences, Ghent University, and funded by the China Scholarship Council (CSC).

GHENT UNIVERSITY

Developing New Approaches for the Analysis of Movement Data

A sport-oriented application

Dissertation submitted in accordance with the requirements for the degree of

Doctor of Science: Geomatics and Surveying

Ontwikkeling van Nieuwe Benaderingen voor de Analyse van Bewegingsgegevens

Een sportgeoriënteerde toepassing

Proefschrift aangeboden tot het behalen van de graad van

Doctor in de Wetenschappen: Geomatica en Landmeetkunde

by/door Pengdong Zhang

Supervisor

Prof. dr. Nico Van de Weghe

Department of Geography, Ghent University

Members of the Examination Committee

Prof. dr. Veerle Van Eetvelde (Chairman)

Department of Geography, Ghent University

Prof. dr. Philippe De Maeyer (Secretary)

Department of Geography, Ghent University

Prof. dr. Yi Qiang

Department of Geography and Environment, University of Hawai'i at Mānoa

Prof. dr. Guy De Tré

Department of Telecommunications and Information Processing, Ghent University

Prof. dr. Peter Bogaert

Department of Geography, Ghent University

Contents

List of Figures

List of Tables

List of Abbreviations

CTM	Continuous Triangular Model
DI	Dynamic Interactions
GIScience	Geographical Information Science
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
MAUP	Modifiable Areal Unit Problem
MSSI	Multi-Scale Straightness Index
MTSSTN	Multi-Temporal Scale Spatio-Temporal Network
MTUP	Modifiable Temporal Unit Problem
OD	Origin-Destination
QTC	Qualitative Trajectory Calculus
REMO	RElative MOtion
RFID	Radio Frequency Identification
RTC	Relative Trajectory Calculus
STC	Space-Time Cube
tevel	Temporal EVolution ELement
TM	Triangular Model

Preface

It is finally the moment to write the preface, which I have been looking forward to for a long time. The past five years pursuing the PhD have been a memorable period in my life. This is a long journey, in which I have experienced so much, including excitement, joyfulness, challenges, frustrations and depressions. With the help from many persons, I eventually persisted till the end of the journey. I would like to take this opportunity to express my appreciation to all of you.

The person I would like to appreciate most is my supervisor Prof. Nico Van de Weghe. Thank you so much for providing me this precious opportunity to pursue the PhD under your supervision. It was so nice an experience working with you. Your scientific vision, attitude, enthusiasm, experience and achievement have influenced me greatly. I appreciate you for your great patience to me during every discussion we have had and the great freedom in dealing with daily work. I am grateful to you for your insightful ideas, comments and advices shared with me on research. I am quite convinced that the many things I have learned from you can certainly influence my work and life positively in the future.

I also would like to express my appreciation to the colleagues at the department. First, I would like to thank all the officemates (former and current) who gave me many help and suggestions on both research and life, and accompanied me during the PhD period. They are Luc Zwartjes, Mathias Versichele, Manuel Claeys Boùùaert, Roel Huybrechts, Tim Baert, Jasper Beernaerts, Alexander Duytschaever and Johannes De Groeve. Other colleagues from the department who offered me help in a variety of forms also deserve my deep appreciation. They are Bart Dewulf, Bart De Wit, Kristien Ooms, Berdien De Roo, Britt Lonneville, Michiel Van den Berghe, Helga Vermeulen, Sofie De Winter, Nathalie Van Nuffel, Karine Van Acker, Sabine Cnudde and Wim Van Roy.

I am also grateful to the Chinese friends I met in Ghent. Your help, company and support made me not feel lonely. They are: Xiuqin Wu, Tao Li, Junzhuo Liu, Pei Lei, Baoyi Yu, Shengrun Zhang, Long Zhang, Gonghuan Fang, Miao Zhang, Jiao Liu, Weiyang Zhang, and many others whom I did not list. Thank you for the memorable moments I had together with you.

I would like to particularly thank my BSc classmate Mengmeng Li for his long-term company since the beginning of my PhD application. Your frequent encouragements, supports and help influenced me during the whole PhD period. I will never forget the many discussions we had on both life and research, and the wonderful moments we had together in both Ghent and Enschede. I would also give my deep appreciation to my MSc classmate Dongliang Peng. We had unforgettable experiences together in both Belgium and Nordic Europe, during which we had many interesting and insightful discussions on research and life.

I would like to give my deepest appreciation to my family members. I appreciate my parents in the deepest heart for their constant understanding and support on me. I am so grateful to my brother and his wife for taking care of our parents and the family so well during the PhD period. Special thanks should give my half-year-old niece for the great happiness she brought to me. Sincerely wish you grow healthily day by day!

At the end, I would like to give my deep and special thanks to Ms. Jiamin Li. Meeting you was so lucky an event in my life, but loving you and being loved by you was even much luckier. Thank you for your continuous trust, understanding, tolerance and company during the past two years. I enjoy and miss the sweet moments we had so much. Look forward to our new chapter in China!

Pengdong Zhang

Ghent, March 2018

1

Introduction

1.1 Background and motivation

With the rapid development of advanced location-aware technologies, such as GPS (global positioning system), RFID (radio frequency identification), Bluetooth, WiFi, image recognition, and video tracking, data related to the trajectories of moving objects can be acquired more easily than ever before. This type of data has different names, such as (geospatial) lifeline data (Laube et al., 2004; Laube et al., 2005), trace data (Pan et al., 2013), trajectory data (Miller and Han, 2009; Zheng, 2015), movement data (Andrienko et al., 2007; Long & Nelson, 2013; Dodge et al., 2016), and mobility data (Giannotti & Pedreschi, 2008; Guo et al., 2012). In this thesis, we mainly adopt the term ‘movement data’ to denote such data. However, in specific cases where ‘movement data’ does not provide a good representation, the terms ‘trajectory’ or ‘trajectory data’ are adopted. The proliferation of rich and voluminous movement data has drawn enormous attention from researchers in various fields, thereby promoting related research topics with respect to movement data. Typical research topics include modelling and simulating movement data (Wang et al., 2016; Yeoman & Duckham, 2016; Ahearn et al. 2017), visual analyses of movement data (Andrienko & Andrienko, 2013), knowledge discovery in movement data (Laube et al., 2005; Wachowicz et al., 2011), similarity measurements of movement data (Dodge et al., 2012; Yuan & Raubal, 2014), scale-aware analyses of movement data (Laube & Purves, 2011; Soleymani et al., 2014), context-aware analyses of movement data (Siła-Nowicka et al., 2016; Sharif & Alesheikh, 2017), dynamic interactions in movement data (Long & Nelson, 2013; Miller, 2015), predictions of movement data (Yavaş et al., 2005; Borkowski, 2017), etc. In addition to the various research topics, the methods and techniques of analysing movement data have already been applied in a large number of domains, such as transportation (Li et al., 2012; Zhang et al., 2017), urban planning (Ratti et al., 2010; Liu et al., 2012), ecology (Long & Nelson, 2013; Miller, 2015), environment (Hsieh et al., 2015), social media (Li et al., 2015; Shi et al., 2016), sport science (Stein et al., 2015; Sacha et al., 2017), and business (Körner et al., 2010). In this thesis, we are particularly motivated by four of the aforementioned research topics (i.e., visual analyses of movement data, knowledge discovery in movement data, scale-aware analyses of movement data, and dynamic interactions in movement data), and we aim to explore their potential applications in a relatively novel domain: sports. In the following, a general overview of these research topics and their current applications for sports is provided.

1.1.1 Visual analysis of movement data

Data visualisation techniques enable us to form multi-dimensional data representations, which can be used to provide knowledge and potential insights into these data (Soukup & Davidson, 2002). The process of data visualisation generally includes data collection and storage, data pre-processing and transformation, and information display and perception (Ware, 2004). Through these steps, interesting information or useful knowledge can be identified. Given the popularity of movement data, the visual analysis of movement data has been an active topic in recent years. Because movement data can represent the trajectories of moving objects in both space and time, these data are also considered spatio-temporal data. Representing spatio-temporal data or time (in particular) along with the two spatial dimensions has been a longstanding problem in geographical information science (GIScience) (Kwan & Neutens, 2014). One of the primary yet still frequently used tools to visually represent such information is the so-called space-time cube (STC), which was originally established in time geography (Hägerstraand 1970). In a STC, spatio-temporal data are visualised in 3D space, where the bottom 2D plane represents the 2D geographical space and the third axis represents time. Since its introduction, STCs have become a popular tool in GIScience to visualise human activity-related data/information, including movement data.

In addition to STCs, a large variety of other tools have also been developed to visually represent movement data from different perspectives. Some of the typical tools are introduced as follows. A traditional tool for the visualisation of movements is flow chart (such as discrete points, flow lines and arrows) which is drawn on a map or image (Vasiliev, 1997). Animation is also a widely used way to visualise movement data, .e.g., the trajectories of objects during a time interval can be visualised by animations (Andrienko et al., 2007; Klein et al., 2014). This is usually effective when the data sets are not large. When the movement data are large, visual displays of all trajectories become unsatisfactory, since problems such as clutter and overprinting can inevitably occur. Thus, aggregation is used for visualising (Andrienko & Andrienko, 2008). Edge bundling is a common method to aggregate movement data for spatial visualisations (Holten et al., 2009; Hoeferlin et al., 2013; Hurter et al., 2014; Takayanagi & Okada, 2015). It is often used for origin-destination (OD) data, i.e., movement data where the start and end points are the main focuses. Kernel density estimation is a frequently used aggregation method as well (Demšar & Virrantaus, 2010; Hurter et al., 2012). It can also be extended to 3D space-time density estimation by combining the STC and the standard 2D kernel density estimation to better visually display movement data. Besides, visualising the attribute information of movement data is important. Cartographical representations have been used for visualise such information. Common tools include flow map (Tobler, 1987; Han et al., 2017) and heatmap (Cakmak et al., 2015). In addition, movement data can also be visualised with linked views to allow incorporation of contextual information (Andrienko et al., 2011; Andrienko & Andrienko, 2013). Other interesting methods/tools can be found in(Randell et al., 1992; Kwan, 2000; Andrienko et al., 2007; Ren & Kwan, 2007; Rinzivillo et al., 2008; Willems et al., 2009; Shamoun-Baranes et al., 2012; Enguehard et al., 2013; Zeng et al., 2013; Wang et al., 2014; Konzack et al., 2017). With the improvements of visual representation tools and techniques, movement data have been better understood over time.

The Triangular Model (TM) is an interesting tool that was originally introduced by Kulpa (Kulpa, 1997; Kulpa, 2006) and has been recently extensively adopted to model and visualise temporal information in a two-dimensional space (Van de Weghe et al., 2007; Qiang et al., 2010; Qiang et al., 2012; Qiang et al., 2012; Qiang et al., 2014). On the basis of the TM, the Continuous Triangular Model (CTM) was then developed by Qiang et al. (2014) to represent temporal information continuously. Currently, the CTM is only preliminarily used as a useful tool for analysing movement data. Qiang et al. (2014) initially used the CTM to simply analyse the movement data of a football match lasting 13 minutes. Subsequently, the CTM was used to visualise repetitive motion patterns in groups of moving objects (Chavoshi et al., 2015). Given the distinctive characteristics of the CTM in representing time-related data (including movement data as well), more extensive explorations on the applications of the CTM in the domain of movement data analysis deserve further attention.

1.1.2 Knowledge discovery in movement data

Data mining has attracted a great deal of attention in both the information industry and society as a whole for a long time because of the wide availability of large amounts of data and the urgent need for converting such data into useful knowledge and information (Han & Kamber, 2006). Hence, data mining is a key technique for gaining knowledge, and a number of data mining methods and algorithms have been developed. Due to the sizable amount of movement data in this data-rich era, movement data mining methods for discovering valuable knowledge that can also be used for various movement data-related purposes are greatly needed.

Knowledge discovery in movement data mainly includes three steps: movement data reconstruction, knowledge extraction and knowledge delivery (Giannotti & Pedreschi, 2008; Dodge, 2011). Movement data reconstruction is considered a pre-processing step, and it mainly includes filtering (e.g., removing outliers), resampling (e.g., obtaining regular sampled data), smoothing (e.g., removing the effect of noise using specific techniques), and map matching for specific data (e.g., matching the position data with the actual map). Knowledge extraction aims to discover patterns and structures in movement data and acquire useful knowledge about the behaviour of moving objects. In this step, movement data mining methods and techniques, such as movement pattern discovery, trajectory classification, trajectory clustering, and movement similarity analysis, are adopted. In the knowledge delivery step, the main purpose is to evaluate or interpret the discovered patterns or other outcomes. To perform evaluations or interpretations, effective visualisation techniques are required so that the results can be appropriately presented and the extracted knowledge can be delivered.

Among the three steps, knowledge extraction using movement data mining methods and techniques plays an important role because different methods may derive different knowledge. With respect to the movement data mining methods, we can broadly divide them into three categories: (1) shape-based methods; (2) attribute-based methods, and (3) shape-and-attribute based methods. Shape-based methods essentially focus on the geometric characteristics (i.e., shape) of the trajectory of a movement because a trajectory can be considered as a series of discrete points in chronological order. Typical shape-based methods include methods that are used for trajectory clustering (Lee et al., 2007; Palma et al., 2008; Zhang et al., 2014), movement pattern mining (Laube et al., 2005; Gudmundsson et al., 2007; Andersson et al., 2008; Jeung et al., 2008; Wachowicz et al., 2011; Jacob & Idicula, 2012; Kjærgaard et al., 2012; Fort et al., 2014; Turdukulov et al., 2014; Loglisci 2017), and trajectory outlier detection (Lee et al., 2008; Yuan et al., 2011; Liu et al., 2012). Attribute-based methods mainly focus on analysing the changes of motion attributes (such as speed, acceleration, distance and direction) that are used to characterise the movements of objects over time. These methods can be used to mine useful patterns (Laube et al., 2005), explore the similarities of trajectories (Dodge et al., 2012; Chavoshi et al., 2015), and even predict the positions of moving objects over time (Sabarish et al., 2015). Shape-and-attribute based methods can be regarded as a mixture of shape-based methods and attribute-based methods. One of the distinct advantages of such methods is that the meanings of trajectories can be enhanced and refined by integrating semantic information (Buchin et al., 2012; Elragal & EL-Gendy, 2013; Buchin et al., 2014).

Given the importance of movement data mining methods and techniques in knowledge discovery in movement data, investigations by researchers from various domains should focus on this important aspect.

1.1.3 Scale-aware analysis of movement data

Scale is an important issue in many disciplines, particularly those that involve space and/or time (e.g., GIScience). Well-known approaches to addressing this issue include the modifiable areal unit problem (MAUP) for the spatial scale and the modifiable temporal unit problem (MTUP) for the temporal scale (Openshaw 1984; Cheng & Adepeju, 2014). In GIScience, scale mostly denotes resolution or extent (Goodchild 2011). Geographical data normally have a specific resolution; thus, operations for such data are scale specific because differences in scale could introduce dramatic changes. Thus, scale is of significant importance, especially after it was included as the fifth dimension in 5D data modelling (van Oosterom & Stoter, 2010). Therefore, scale must be considered when analysing space and/or time-related data.

For movement data, relatively few researches have been undertaken from a cross-scale or multi-scale perspective. Laube & Purves (2011) investigated the changes of three motion attributes (i.e., speed, turning angle and sinuosity) of ten cows at six keenly selected temporal scales. Long & Nelson (2013) argued that dynamic interactions can be analysed from four analysis levels of scale: local, interval, episodal and global. Postlethwaite et al. (2013) presented the multi-scale measure MSSI (multi-scale straightness index) to analyse animal movement data at multiple temporal scales. Soleymani et al. (2014) proposed a methodology to explore the behavioural movement of zebrafish by joint spatio-temporal cross-scale analyses of three motion attributes (i.e., speed, acceleration and sinuosity). In addition, a continuous spatio-temporal model (CSTM) was proposed by Van de Weghe et al. (2014). The CSTM is a conceptual model that integrates both space and time over multiple scales. In theory, the CSTM has a strong ability to analyse movement data from either a spatial, temporal or spatio-temporal perspective at multiple levels. However, this is currently only a theoretical model. Real-world applications still need to be extended in the future. A disadvantage of these research approaches is that apparent discrepancies between the findings of the research and the real circumstances might exist because only a very small number of carefully selected temporal scales have been considered in the analyses. More precise findings can be obtained with additional temporal scales.

1.1.4 Dynamic interactions in movement data

Moving objects commonly move in geographical space, in which the geographical context (e.g., the environments where moving objects live) is considered one of the important components. Therefore, movement data interactions can be categorised as the interactions between geographical contexts (Gray & Moseley, 2005; Seneviratne et al., 2010), between moving objects and geographical contexts, and between the moving objects themselves. Note that as the first two can be considered as a context-aware analysis of movement data, they are not within the scope of this thesis. In this thesis, we only consider the interactions between/among the moving objects themselves. Interactions can be classified as static interactions or dynamic interactions (Doncaster, 1990). In movement data, static interactions are purely described by spatial properties (without taking into account the possibility of temporal avoidance or attraction between individuals), whereas dynamic interactions are defined based on both spatial and temporal components (Miller, 2015). In this thesis, the main focus is on dynamic interactions in movement data.

In general, dynamic interactions can be defined as the way the movements of individuals are related or the inter-dependency among the movements of individuals. For example, attraction and avoidance are two typical types of dynamic interactions (Miller, 2015). The research on dynamic interactions in movement data is still in its infancy. In GIScience, typical research related to movement data is listed as follows. Miller (2012) analysed the dynamic interactions between individuals based on the GPS data of animals using five different techniques, thereby comparing the results acquired by the different techniques. Subsequently, a null model approach (Miller, 2015) was developed by the same author to compare six dynamic interaction metrics. The approach was tested based on the data of five brown hyena dyads in northern Botswana. The comparison results highlight the need for further studies to identify appropriate methods for measuring and interpreting dynamic interactions (Miller, 2015). Long & Nelson (2013) introduced the method Dynamic Interactions (DI) for measuring the dynamic interactions between pairs of moving objects. The method was validated on six simulated datasets and two applied examples (i.e., team sports and wildlife). The results showed that the DI method can be used to measure dynamic interactions in movement data. Long et al. (2014) executed an examination of eight currently available indices of dynamic interactions in wildlife telemetry studies and compared the effectiveness of the indices. In Long (2015), the statistical properties of a suite of currently available methods for evaluating dynamic interactions were examined. Konzack et al. (2017) proposed a new approach to analyse interactions between two trajectories and developed a prototype visual analytics tool to evaluate the approach based on three datasets.

In all, the aforementioned research mainly focuses on either comparing/evaluating existing methods with respect to dynamic interactions based on various datasets or developing new methods of measuring dynamic interactions between two moving objects. Most of the research has been undertaken at a single temporal scale. In addition, few studies have focused on exploring the importance of each moving object and identifying moving objects that play relatively important roles in maintaining specific interactions. Hence, new efforts still have to be made to extend the research.

1.1.5 Applications of movement data analysis in sports

Recent developments in sensor or location-aware techniques have resulted in an increasing interest in recording and analysing movement in team sports. A team sport includes any sport that involves two or more players working together towards a shared objective, and it can be considered group movement in which individuals collaborate and compete following specific rules (Stein et al., 2017). Typical team sports include soccer/football, basketball, hockey, baseball, etc. Traditional analyses in team sports mainly rely on descriptive statistics of the data obtained. These approaches typically focus on basic analyses, such as distance, speed, sprints, heat maps, and preferred attacking side (Feuerhake 2016). However, traditional analysis methods are not well suited for more advanced tasks in team sports, such as passing possibilities, pass sequence patterns, and (frequent) movement patterns (Feuerhake 2016; Gudmundsson & Horton, 2017). Benefiting from the popularity of sports movement data in recent years and the various methods/techniques in movement data analysis, more advanced analyses can be executed, thereby facilitating the evaluation of player performance for sports professionals (e.g., coaches) during matches or training (Stein et al., 2017).

Therefore, a series of research work with respect to sports movement data has been undertaken in recent years. Among the many investigated sports, football has received much attention. The typical investigations focused on football are introduced below. Iwase & Saito (2002) proposed a method of tracking the positions of a specific football player by using multiple cameras. Subsequently, the authors extended this method to enable tracking the positions of multiple players (Iwase & Saito, 2003). In addition to the tracking of players, methods to track the ball were also developed (Ren et al., 2009). Thus, the trajectories of either the players or the balls can be acquired, and various advanced analyses have been executed based on trajectories. Laube et al. (2005) presented a generic geographical knowledge discovery approach called Relative Motion (REMO) to explore the movements of objects, and they applied the approach to a movement dataset consisting of 11 football players lasting 33 seconds. Kim et al. (2011) introduced a framework for the tactical analysis of football matches based on the trajectories of both the players and the ball using spatial and spatio-temporal approaches. Lucey et al. (2013) presented an approach to assessing team strategies based on the tracking data from the English Premier League (2010-2011 season) and investigated possible reasons why the home advantage exists in football. Gudmundsson & Wolle (2014) developed a collection of spatio-temporal tools specifically for the performance analysis of football players and teams. Horton et al. (2015) presented a model to automatically classify the passing in football. Feuerhake (2016) presented an approach for the recognition of movement patterns as an advanced analysis based on the players' trajectories. Sacha et al. (2017) proposed a novel dynamic approach for abstracting players' trajectories by combining trajectory simplification and clustering techniques to support the interpretation and understanding of movement patterns. In addition to football, movement data are used in other sports for related research as well. Demaj (2013) presented an approach for post-match analysis based on the trajectories of tennis balls and visualised the spatio-temporal patterns. Lucey et al. (2014) investigated how to obtain an open shot by analysing team movements in basketball based on the tracking data. Chavoshi et al. (2015) proposed a novel approach for measuring the similarity in the interactions between moving objects and validated the approach using the movement data tracked from samba and tango dancers. Metulini et al. (2017) presented a method of identifying homogeneous spatial relations among players in the court, differentiating defensive or offensive actions, and analysing the transition probabilities from a certain group to another.

The aforementioned research shows that with the development of methods and techniques for the analysis of movement data, the subject of sports analytics has stepped into a new stage where more advanced analysis tasks can be achieved and thus more potential insights can be provided to sports professionals (e.g., coaches) for suitable tactical arrangements. Moreover, the demand for sports analytics may represent driving forces for the emergence of new methods and techniques of analysing movement data. Hence, movement data analysis has the potential for use in a number of applications for sports in the future.

1.2 Research objectives

The previous section presented an overview of the state-of-the-art research of four topics with respect to the analysis of movement data and the application of movement data analysis in sports. Based on this overview, the thesis presented here proposes four general research questions (RQs), which are addressed in this section.

RQ 1: Can the CTM bring added value to the analysis of movement data?

This research question focuses on the applications of the CTM in the analysis of movement data based on its functionalities, and it is addressed in Chapters 2, 3 and 4 from different perspectives.

Chapter 2 addresses an exhaustive application of the CTM in the visual analysis of football movement data obtained from a real and entire football match. In this chapter, three representative motion attributes (i.e., speed, ball possession and territorial advantage) that are either common for all types of moving objects or particularly meaningful for specific types of moving objects are visualised using the CTM. Based on the generated CTM diagrams, the performance of either the players or the whole team can be explored in detail; thus, insightful information can be generated for sports professionals. Moreover, the implementation of algebra operators (e.g., summation, subtraction, maximum, minimum, and mean) allows the CTM to generate more advanced diagrams that can be used to address more sophisticated tasks. Based on the advanced functionalities, we propose a new measure for deeply exploring the performance of the whole team and visualising it using the CTM. With the help of the CTM diagrams, interesting information that is difficult to obtain using other tools can be obtained easily; thus, more abundant information can be provided for tactical improvements or arrangements. This chapter shows the strong ability of the CTM in the analysis of movement data based on its visualisation functionalities. Because of its strong visual analysis capabilities, the CTM is used as a visualisation tool as well in Chapter 4.

In Chapter 4, the main focus is the exploration of dynamic interactions in movement data. In this chapter, a hybrid approach that combines the multi-temporal scale spatio-temporal network (MTSSTN) and the CTM is proposed. Based on the proposed approach, the interaction intensities between any two individual objects or among multiple objects can be explored by visualising the corresponding interaction intensity measures using the CTM. By using the extended functionalities of the CTM, we can further generate more advanced CTM diagrams. Based on the visualisations, the importance of each individual can be explored and the most important individuals can be identified. The proposed approach is validated using part of the aforementioned football movement data, and the results demonstrate the effectiveness of the approach in the application of football.

In addition to its visualisation capabilities, the CTM has a strong ability to represent temporal data via multiple temporal scales. The results demonstrated in Chapters 2 and 4 are also analysed from a multi-scale perspective. In addition to the brief introduction of the multi-scale ability of the CTM, we address this feature in more detail in Chapter 3. Chapter 3 mainly focuses on the topic of knowledge discovery in movement data using a novel cross-scale oriented sequence analysis approach. The key to the proposed approach is the construction of sequences based on the CTM. Based on the different scale properties, four types of sequences are derived. Part of the aforementioned football movement data is used to validate the effectiveness of the proposed approach. The findings demonstrate that the proposed approach is useful and efficient in mining information from movement data, which shows that the CTM can also be applied for knowledge discovery using movement data.

RQ 2: What interesting information can be discovered in football movement data?

This research question is addressed in Chapters 3 and 5. In Chapter 3, we propose a cross-scale oriented sequence analysis approach. Based on the proposed approach, two distinct research aims are derived: investigating the changes in motion attributes of players across different temporal scales and detecting the time intervals during which active events might have occurred. We then apply this approach to football movement data. The results indeed reveal interesting information on the changes in motion attributes of the players across different temporal scales, which has been rarely obtained by former research and thus might be helpful for coaches to arrange tactics more scientifically. In addition, by automatically detecting the time intervals during which active events might have occurred, coaches can quickly adjust to specific time intervals for a post-analysis of the match rather than watching the video second by second to locate target time intervals. The findings indeed appear interesting from different perspectives.

Chapter 5 proposes a novel Reeb graph-based approach to discover moving flock patterns formed by the players in the same team. Although considerable work on methods of identifying moving flock patterns has been performed for movement data in the field of GIScience, few researchers have developed such methods for football. In Chapter 5, we first develop an improved definition of moving flock and then propose a taxonomy of moving flock patterns, which are used to derive eight types of interesting moving flock patterns. Finally, the approach is applied to part of the football movement data. The results demonstrate that the proposed approach is capable of differentiating various types of moving flock patterns. Based on the discovered moving flock patterns, coaches can identify players that can form different types of groups, which can increase the confidence associated with arranging corresponding tactics according to specific demands.

RQ 3: Can added values be provided if taking multiple (temporal) scales into account when analysing movement data?

Chapters 3 and 4 consider the temporal scale when developing corresponding approaches for analysing movement data. In Chapter 3, a cross-scale oriented sequence analysis approach is proposed and then applied to part of the football movement data with a duration of 95 minutes. When analysing the football movement data using the proposed approach, 96 temporal scales are considered, which is a comparatively large number of temporal scales. The findings reveal the rules of changes for motion attributes across different temporal scales, which can be used to provide more abundant information to sports professionals. In addition, when detecting the time intervals during which active events might have occurred, the results based on the approach that involves multiple temporal scales are better than those involving only one temporal scale.

In Chapter 4, a hybrid approach combining the MTSSTN and the CTM is proposed to explore the dynamic interactions in movement data. One distinctive characteristic of this approach is the generation of the MTSSTN, which to our knowledge is the first spatio-temporal network that considers the multi-temporal scale. The results based on the MTSSTN are visualised using the CTM, and information at all temporal scales is displayed. Hence, based on the generated CTM diagrams, much more detailed information can be revealed compared with those that only use one temporal scale. The two chapters demonstrate that added values can be provided if multiple (temporal) scales are considered when analysing movement data.

RQ 4: What efforts can be contributed to the relatively new research topic of dynamic interactions in movement data?

As previously stated, the research on dynamic interactions in movement data is relatively new compared with other research topics with respect to the analysis of movement data. In this thesis, we contribute efforts to the exploration of dynamic interactions in movement data, which is addressed in detail in Chapter 4. In this chapter, we mainly aim to quantitatively explore the interaction intensities between two individual objects or among multiple objects. This investigation is achieved based on a specific interaction pattern. In all, three types of interaction patterns are derived based on the Relative Trajectory Calculus (RTC) developed by Van de Weghe (2014). One distinctive characteristic with respect to previous research is that we propose a method of exploring the dynamic interactions among multiple individuals. We also propose a method of exploring the importance of each individual in maintaining the dynamic interactions and identifying the individuals who play important roles in maintaining each interaction pattern. Part of the aforementioned football movement data is used as a case study to validate the effectiveness of the proposed approach. The results demonstrate that the proposed approach is useful for exploring dynamic interactions in movement data and discovering insightful information. We are convinced that this approach represents an important contribution to the research topic and will propel its development in the future.

1.3 Thesis outline

This thesis consists of six chapters. Chapters 2 to 5 are the substantial parts of this thesis and include four academic articles that have been published by or submitted to international peer-reviewed journals or will be submitted in the future. Each chapter of this thesis is organised to answer the previously mentioned research questions. Because certain chapters may answer part of the research questions while others might respond to a specific research question, the chapters are not grouped. Strong links are observed between the chapters presented in this thesis. To ensure that these articles can be read smoothly and independently from each other, inevitable overlaps are included in the individual chapters with regard to the literature reviews and the description of the basic concept (e.g., the CTM) and dataset used in this thesis. Chapter 6 summarises the main findings and contributions of this thesis and proposes avenues for future research. The general outline of this thesis is shown in Figure 1.1.

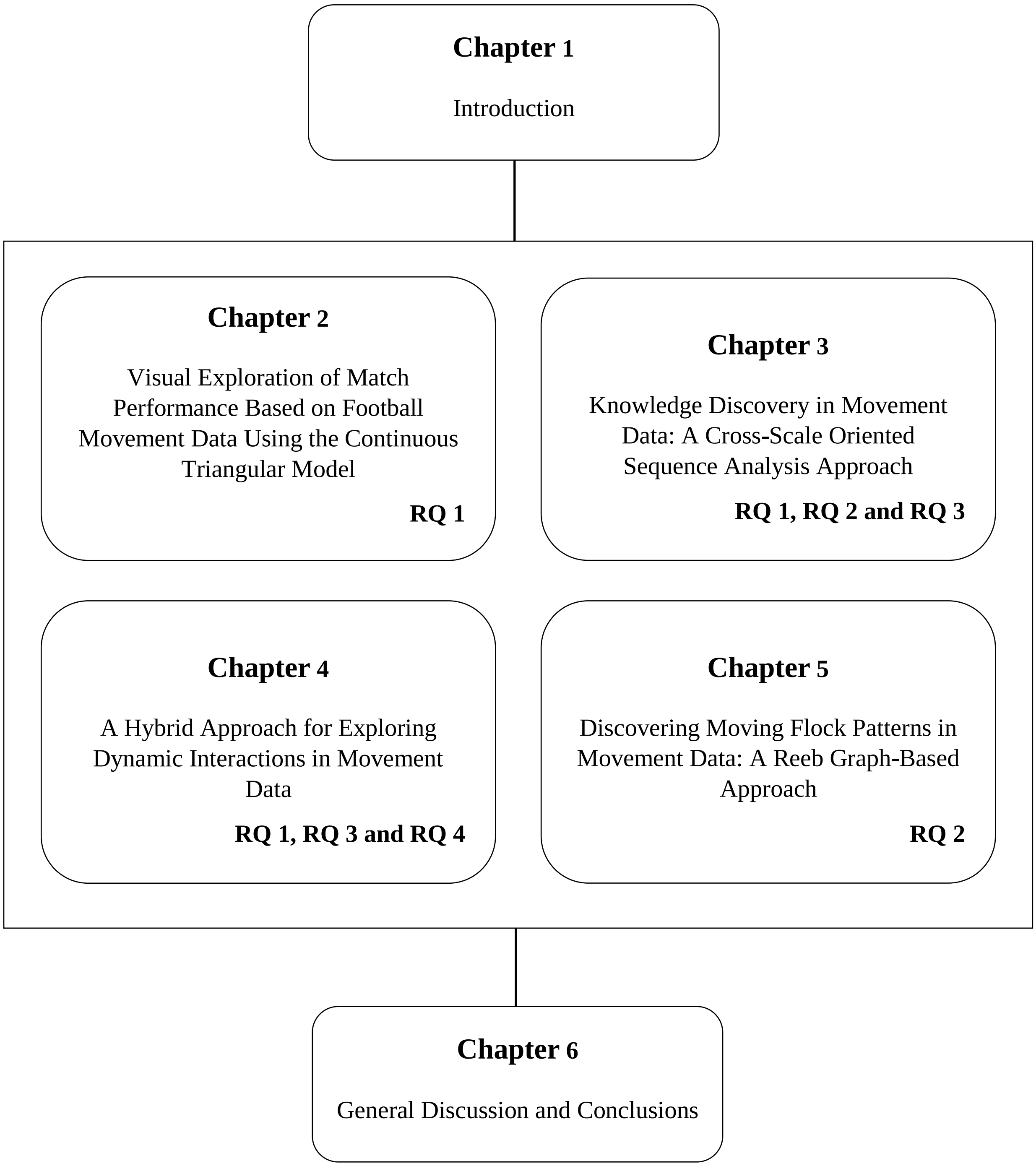


Figure 1.1. Thesis outline.

References

Ahearn, S. C., Dodge, S., Simcharoen, A., Xavier, G., & Smith, J. L. (2017). A context-sensitive correlated random walk: a new simulation model for movement. *International Journal of Geographical Information Science, 31*(5), 867-883.

Andersson, M., Gudmundsson, J., Laube, P., & Wolle, T. (2008). Reporting leaders and followers among trajectories of moving point objects. *GeoInformatica*, *12*(4), 497-528.

Andrienko, G., Andrienko, N., & Wrobel, S. (2007). Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter*, *9*(2), 38-46.

Andrienko, G., & Andrienko, N. (2008). Spatio-temporal aggregation for visual analysis of movements. In *Proceedings of IEEE Symposium on Visual Analytics Science and Technology* (pp. 51-58).

Andrienko, G., Andrienko, N., & Heurich, M. (2011). An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science*, *25*(9), 1347-1370.

Andrienko, N., & Andrienko, G. (2013). A visual analytics framework for spatiotemporal analysis and modelling. *Data Mining and Knowledge Discovery*, *27*, 55-83.

Borkowski, P. (2017). The ship movement trajectory prediction algorithm using navigational data fusion. *Sensors*, *17*(6), 1432.

Buchin, M., Dodge, S., & Speckmann, B. (2012). Context-aware similarity of trajectories. *Lecture Notes in Computer Science*, *7478*, 43-56.

Buchin, M., Dodge, S., & Speckmann, B. (2014). Similarity of trajectories taking into account geographic context. *Journal of Spatial Information Science*, *9*, 101-124.

Cakmak, E., Gärtner, A., Hepp, T., Buchmüller, J., Fischer, F., & Keim, D. A. (2015). Applying visual analytics to explore and analyze movement data. In *Proceedings of the 2015 IEEE Conference on Visual Analytics Science and Technology* (pp. 127-128).

Chavoshi, S. H., De Baets, B., Neutens, T., De Tré, G., & Van de Weghe, N. (2015). Exploring dace movement data using sequence alignment methods. *PloS one*, *10*(7), e0132452.

Chavoshi, S. H., De Baets, B., Qiang, Y., De Tré, G., Neutens, T., & Van de Weghe, N. (2015). A qualitative approach to the identification, visualisation and interpretation of repetitive motion patterns in groups of moving point objects. *International Arab Journal of Information Technology*, *12*(5), 415-423.

Cheng, T., & Adepeju, M. (2014). Modifiable temporal unit problem (MTUP) and its effect on space-time cluster detection. *PloS one*, *9*(6), e100465.

Demaj, D. (2013). Geovisualizing spatio-temporal patterns in tennis: an alternative approach to post-match analysis. In *Proceedings of the 26th International Cartographic Conference*.

Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Van de Weghe, N., ... & Weibel, R. (2015). Analysis and visualisation of movement: an interdisciplinary review. *Movement ecology*, *3*(1), 5.

Demšar, U., & Virrantaus, K. (2010). Space–time density of trajectories: exploring spatio-temporal patterns in movement data. *International Journal of Geographical Information Science*, *24*(10), 1527-1542.

Dodge, S. (2011). *Exploring movement using similarity analysis*. University of Zurich, Zurich.

Dodge, S, Laube, P., & Weibel, R. (2012). Movement similarity assessment using symbolic representation of trajectories. *International Journal of Geographical Information Science*, *26*(9), 1563-1588.

Dodge, S., Weibel, R., Ahearn, S. C., Buchin, M., & Miller, J. A. (2016). Analysis of movement data. *International Journal of Geographical Information Science*, *30*(5), 825-834.

Doncaster, C. P. (1990). Non-parametric estimates of interaction from radio-tracking data. *Journal of Theoretical Biology*, *143*(4), 431-443.

Elragal, A., & EL-Gendy, N. (2013). Trajectory data mining: integrating semantics. *Journal of Enterprise Information Management*, *26*(5), 516-535.

Enguehard, R. A., Hoeber, O., & Devillers, R. (2013). Interactive exploration of movement data: a case study of geovisual analytics for fishing vessel analysis. *Information visualization*, *12*(1), 65-84.

Feuerhake, U. (2016). Recognition of repetitive movement patterns-the case of football analysis. *ISPRS International Journal of Geo-Information*, *5*(11), 208.

Fort, M., Sellarès, J. A., & Valladares, N. (2014). A parallel GPU-based approach for reporting flock patterns. *International Journal of Geographical Information Science*, *28*(9), 1877-1903.

Giannotti, G., & Pedreschi, D. (2008). *Mobility, data mining and privacy: geographic knowledge discovery*. Berlin Heidelberg: Springer-Verlag.

Goodchild, M. F. (2011). Scale in GIS: an overview. *Geomorphology*, *130*(1), 5-9.

Gray, L. C., & Moseley, W. G. (2005). A geographical perspective on poverty–environment interactions. *The Geographical Journal*, *171*(1), 9-23.

Gudmundsson, J., van Kreveld, M., & Speckmann, B. (2007). Efficient detection of patterns in 2D trajectories of moving points. *GeoInformatica*, *11*(2), 195-215.

Gudmundsson, J., & Horton, M. (2017). Spatio-temporal analysis of team sports. *ACM Computing Surveys*, *50*(2), 22.

Gudmundsson, J., & Wolle, T. (2014). Football analysis using spatio-temporal tools. *Computers, Environment and Urban Systems*, *47*, 16-27.

Guo, D., Zhu, X., Jin, H., Gao, P., & Andris, C. (2012). Discovering spatial patterns in origin-destination mobility data. *Transactions in GIS*, *16*(3), 411-429.

Hägerstraand, T. (1970). What about people in regional science? *Papers in Regional Science*, *24*(1), 7-24.

Han, J., & Kamber, M. (2006). *Data mining: concepts and techniques*. Morgan kaufmann.

Han, S. Y., Clarke, K. C., & Tsou, M. H. (2017). Animated flow maps for visualizing human movement: two demonstrations with air traffic and twitter data. In *Proceedings of the 1st ACM SIGSPATIAL Workshop on Analytics for Local Events and News* (pp. 5).

Hoeferlin, M., Hoeferlin, B., Heidemann, G., & Weiskopf, D. (2013). Interactive schematic summaries for faceted exploration of surveillance video. *IEEE transactions on* *multimedia*, *15*(4), 908-920.

Holten, D., & Van Wijk, J. J. (2009). Force-directed edge bundling for graph visualization. *Computer Graphics Forum*, *28*(3), 983-990.

Horton, M., Gudmundsson, J., Chawla, S., & Estephan, J. (2015). Automated classification of passing in football. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 319-330).

Hsieh, H. P., Lin, S. D., & Zheng, Y. (2015). Inferring air quality for station location recommendation based on urban big data. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 437-446).

Hurter, C., Ersoy, O., Fabrikant, S. I., Klein, T. R., & Telea, A. C. (2014). Bundled visualization of dynamic graph and trail data. *IEEE transactions on visualization and computer graphics*, *20*(8), 1141-1157.

Hurter, C., Ersoy, O., & Telea, A. (2012). Graph bundling by kernel density estimation. *Computer Graphics Forum*, *31*(3), 865-874.

Iwase, S., & Saito, H. (2002). Tracking soccer player using multiple views. In *MVA* (pp. 102-105).

Iwase, S., & Saito, H. (2003). Tracking soccer players based on homography among multiple views. In *Proceedings of Visual Communications and Image* (pp. 283-292).

Jacob, G. M., & Idicula, S. M. (2012). Detection of flock movement in spatio-temporal database using clustering techniques-An experience. In *Proceedings of the 2012 IEEE International Conference on Data Science & Engineering* (pp. 69-74).

Jeung, H., Yiu, M. L., Zhou, X., Jensen, C. S., & Shen, H. T. (2008). Discovery of convoys in trajectory databases. In *Proceedings of the VLDB Endowment* (pp. 1068-1080).

Kim, H., Kwon, O., & Li, K. (2011). Spatial and spatiotemporal analysis of soccer. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 385-388).

Kjærgaard, M. B., Wirz, M., Roggen, D., & Tröster, G. (2012). Mobile sensing of pedestrian flocks in indoor environments using wifi signals. In *Proceedings of the 2012 IEEE International Conference on Pervasive Computing and Communications* (pp.95-102).

Klein, T., Van Der Zwan, M., & Telea, A. (2014). Dynamic multiscale visualization of flight data. In *2014 IEEE International Conference on Computer Vision Theory and Applications* (pp. 104-114).

Konzack, M., McKetterick, T., Ophelders, T., Buchin, M., Giuggioli, L., Long, J., ... & Buchin, K. (2017). Visual analytics of delays and interaction in movement data. *International Journal of Geographical Information Science*, *31*(2), 320-345.

Körner, C., Hecker, D., May, M., & Wrobel, S. (2010). Visit potential: a common vocabulary for the analysis of entity-location interactions in mobility applications. In *Geospatial Thinking* (pp. 79-95).

Kulpa, Z. (1997). Diagrammatic representation of interval space in proving theorems about interval relations. *Reliable Computing*, *3*(3), 209-217.

Kulpa, Z. (2006). A diagrammatic approach to investigate interval relations. *Journal of Visual Languages and Computing*, *17*(5), 466-502.

Kwan, M. P. (2000). Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. *Transportation Research Part C: Emerging Technologies*, *8*(1), 185-203.

Kwan, M. P., & Neutens, T. (2014). Space-time research in GIScience. *International Journal of Geographical Information Science*, *28*(5), 851-854.

Laube, P., Imfeld, S., & Weibel, R. (2005). Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science*, *19*(6), 639-668.

Laube, P., van Kreveld, M., & Imfeld, S. (2005). Finding REMO – detecting relative motion patterns in geospatial lifelines. In *Proceedings of the 11th International Symposium on Spatial Data Handling* (pp. 201-215).

Laube, P., & Purves, R. S. (2011). How fast is a cow? Cross-scale analysis of movement data. *Transactions in GIS*, *15*(3), 401-418.

Lee, J. G., Han, J., & Li, X. (2008). Trajectory outlier detection: a partition-and-detect framework. In *Proceedings of the International Conference on Data Mining* (pp. 140-149).

Lee, J. G., Han, J., & Whang, K. (2007). Trajectory clustering, a partition-and-group framework. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data* (pp. 593-604).

Li, J., Qin, Q., Han, J., Tang, L. A., & Lei, K. H. (2015). Mining trajectory data and geotagged data in social media for road map inference. *Transactions in GIS*, *19*(1), 1-18.

Li, X., Pan, G., Wu, Z., Qi, G., Li, S., Zhang, D., ... & Wang, Z. (2012). Prediction of urban human mobility using large-scale taxi traces and its applications. *Frontiers of Computer Science*, *6*(1), 111-121.

Liu, L., Qiao, S., Zhang, Y., & Hu, J. (2012). An efficient outlying trajectories mining approach based on relative distance. *International Journal of Geographical Information Science*, *26*(10), 1789-1810.

Liu, Y., Wang, F., Xiao, Y., & Gao, S. (2012). Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning*, *106*(1), 73-87.

Loglisci, C. (2017). Using interactions and dynamics for mining groups of moving objects from trajectory data. *International Journal of Geographical Information Science*, 1-33.

Long, J. A. (2015). Quantifying spatial-temporal interactions from wildlife tracking data: issues of space, time, and statistical significance. *Procedia Environmental Sciences*, *26*, 3-10.

Long, J. A., Nelson, T. A., Webb, S. L., & Gee, K. L. (2014). A critical examination of indices of dynamic interaction for wildlife telemetry studies. *Journal of Animal Ecology*, *83*(5), 1216-1233.

Long, J. A., & Nelson, T. A. (2013). A review of quantitative methods for movement data. *International Journal of Geographical Information Science*, *27*(2), 292–318.

Long, J. A., & Nelson, T. A. (2013). Measuring dynamic interaction in movement data.

Transactions in GIS, 17(1), 62-77.

Lucey, P., Bialkowski, A., Carr, P., Yue, Y., & Matthews, I. (2014). How to get an open shot: analyzing team movement in basketball using tracking data. In *Proceedings of the 8th annual MIT SLOAN sports analytics conference*.

Lucey, P., Oliver, D., Carr, P., Roth, J., & Matthews, I. (2013). Assessing team strategy using spatiotemporal data. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1366-1374).

Metulini, R., Manisera, M., & Zuccolotto, P. (2017). Space-time analysis of movements in basketball using sensor data. *arXiv preprint*, 1707.00883.

Miller, J. A. (2015). Towards a better understanding of dynamic interaction metrics for wildlife: a null model approach. *Transactions in GIS, 19*(3), 342-361.

Miller, H. J., & Han, J. (2009). *Geographic data mining and knowledge discovery*. CRC Press.

Miller, J. A. (2012). Using spatially explicit simulated data to analyze animal interactions: a case study with brown hyenas in Northern Botswana. *Transactions in GIS*, *16*(3), 271-291.

Openshaw, S. (1984). *The modifiable areal unit problem*. University of East Anglia.

Palma, A. T., Bogorny, V., Kuijpers, B., & Alvares, L. O. (2008). A clustering-based approach for discovering interesting places in trajectories. In *Proceedings of the 2008 ACM symposium on Applied computing* (pp. 863-868).

Pan, G., Qi, G., Zhang, W., Li, S., Wu, Z., & Yang, L. T. (2013). Trace analysis and mining for smart cities: issues, methods, and applications. *IEEE Communications Magazine*, *51*(6), 120-126.

Postlethwaite, C. M., Brown, P., & Dennis, T. E. (2013). A new multi-scale measure for analyzing animal movement data. *Journal of Theoretical Biology*, *317*, 175-185.

Qiang, Y., Chavoshi, S. H., Logghe, S., De Maeyer, P., & Van de Weghe, N. (2014). Multi-scale analysis of linear data in a two-dimensional space. *Information Visualization*, *13*(3), 248-265.

Randell, D. A., Cui, Z., & Cohn, A. G. (1992). A spatial logic based on regions and connection. In *Proceedings of the 3rd International Conference on Knowledge Representation and Reasoning* (pp. 165-176).

Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., ... & Strogatz, S. H. (2010). Redrawing the map of Great Britain from a network of human interactions. *PloS one*, *5*(12), e14248.

Ren, F., & Kwan, M. P. (2007). Geovisualization of human hybrid activity-travel patterns. *Transactions in GIS*, *11*(5), 721-744.

Ren, J., Orwell, J., Jones, G. A., & Xu, M. (2009). Tracking the soccer ball using multiple fixed cameras. *Computer Vision and Image Understanding*, *113*(5), 633-642.

Rinzivillo, S., Pedreschi, D., Nanni, M., Giannotti, F., Andrienko, N., & Andrienko, G. (2008). Visually driven analysis of movement data by progressive clustering. *Information Visualization*, *7*(3-4), 225-239.

Sabarish, B. A., Karthi, R., & Gireeshkumar, T. (2015). A survey of location prediction using trajectory mining. In *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems* (pp. 119-127).

Sacha, D., Al-Masoudi, F., Stein, M., Schreck, T., Keim, D. A., Andrienko, G., & Janetzko, H. (2017). Dynamic visual abstraction of soccer movement. *Computer Graphics Forum*, *36*(3), 305-315.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., ... & Teuling, A. J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3-4), 125-161.

Shamoun-Baranes, J., van Loon, E. E., Purves, R. S., Speckmann, B., Weiskopf, D., & Camphuysen, C. J. (2012). Analysis and visualization of animal movement. *Biology Letters*, *8*(1), 6-9.

Sharif, M., & Alesheikh, A. A. (2017). Context-awareness in similarity measures and pattern discoveries of trajectories: a context-based dynamic time warping method. *GIScience & Remote Sensing*, *54*(3), 426-452.

Shi, Y., Deng, M., Yang, X., Liu, Q., Zhao, L., & Lu, C. T. (2016). A framework for discovering evolving domain related spatio-temporal patterns in Twitter. *ISPRS*

International Journal of Geo-Information, *5*(10), 193.

Siła-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demšar, U., & Fotheringham, A. S. (2016). Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science*, *30*(5), 881-906.

Soleymani, A., Cachat, J., Robinson, K., Dodge, S., Kalueff, A., & Weibel, R. (2014). Integrating cross-scale analysis in the spatial and temporal domains for classification of behavioral movement. *Journal of Spatial Information Science*, *8*(8), 1-25.

Soukup, T., & Davidson, I. (2002). *Visual data mining: Techniques and tools for data visualization and mining*. John Wiley & Sons.

Stein, M., Häußler, J., Jäckle, D., Janetzko, H., Schreck, T., & Keim, D. A. (2015). Visual soccer analytics: understanding the characteristics of collective team movement based on feature-driven analysis and abstraction. *ISPRS International Journal of Geo-Information*, *4*(4), 2159-2184.

Stein, M., Janetzko, H., Seebacher, D., Jäger, A., Nagel, M., Hölsch, J., ... & Grossniklaus, M. (2017). How to make sense of team sport data: from acquisition to data modeling and research aspects. *Data*, *2*(1), 2.

Takayanagi, R., & Okada, Y. (2015). Visualization system by combinatorial use of edge bundling and treemap for network traffic data analysis. In *International Workshop on Information Search, Integration, and Personalization* (pp. 128-141).

Tobler, W. R. (1987). Experiments in migration mapping by computer. *The American Cartographer*, *14*(2), 155-163.

Turdukulov, U., Calderon Romero, A. O., Huisman, O., & Retsios, V. (2014). Visual mining of moving flock patterns in large spatio-temporal data sets using a frequent pattern approach. *International Journal of Geographical Information Science*, *28*(10), 2013-2029.

Van de Weghe, N. (2004). *Representing and reasoning about moving objects: a qualitative approach*. Ghent University.

Van de Weghe, N., De Roo, B., Qiang, Y., Versichele, M., Neutens, T., & De Maeyer, P.

(2014). The continuous spatio-temporal model (CSTM) as an exhaustive framework for multi-scale spatio-temporal analysis. *International Journal of Geographical Information Science*, *28*(5), 1047-1060.

Van de Weghe, N., Docter, R., De Maeyer, P., Bechtold, B., & Rychbosch, K. (2007). The Triangular Model as an instrument for visualising and analysing residuality. *Journal of Archaeological Science*, *34*(4), 649-655.

Vasiliev, I. (1997). Mapping time. *Cartographica: The international journal for geographic information and geovisualization*, *34*(2), 1-51.

van Oosterom, P., & Stoter, J. (2010). 5D data modelling: full integration of 2D/3D space, time and scale dimensions. *Lecture Notes in Computer Science*, *6292*, 310-324.

Wachowicz, M., Ong, R., Renso, C., & Nanni, M. (2011). Finding moving flock patterns among pedestrians through collective coherence. *International Journal of Geographical Information Science*, *25*(11), 1849-1864.

Wang, J., Duckham, M., & Worboys, M. (2016). A framework for models of movement in geographic space. *International Journal of Geographical Information Science*, *30*(5), 970-992.

Wang, Z., Ye, T., Lu, M., Yuan, X., Qu, H., Yuan, J., & Wu, Q. (2014). Visual exploration of sparse traffic trajectory data. *IEEE transactions on visualization and computer graphics*, *20*(12), 1813-1822.

Ware, C. (2004). *Information Visualization: Perception for Design* (2nd ed.). Elsevier.

Willems, N., Van De Wetering, H., & Van Wijk, J. J. (2009). Visualization of vessel movements. *Computer Graphics Forum*, *28*(3), 959-966.

Yavaş, G., Katsaros, D., Ulusoy, Ö., & Manolopoulos, Y. (2005). A data mining approach for location prediction in mobile environments. *Data & Knowledge Engineering*, *54*(2), 121-146.

Yeoman, J., & Duckham, M. (2016). Decentralized detection and monitoring of convoy patterns. *International Journal of Geographical Information Science*, *30*(5), 993-1011.

Yuan, G., Xia, S., Zhang, L., Zhou, Y., & Ji, C. (2011). Trajectory outlier detection algorithm based on structural features. *Journal of Computational Information Systems*, *7*(11), 4137-4144.

Yuan, Y., & Raubal, M. (2014). Measuring similarity of mobile phone user trajectories–a spatio-temporal edit distance method. *International Journal of Geographical Information Science*, *28*(3), 496-520.

Zeng, W., Fu, C, Arisona, S. M., & Qu, H. (2013). Visualising interchange patterns in massive movement data. *Computer Graphics Forum*, *32*(3), 271-280.

Zhang, P., Deng, M., Shi, Y., & Zhao, L. (2017). Detecting hotspots of urban residents' behaviours based on spatio-temporal clustering techniques. *GeoJournal*, *82*(5), 923-935.

Zhang, P., Deng, M., & Van de Weghe, N. (2014). Clustering spatio-temporal trajectories based on kernel density estimation. *Lecture Notes in Computer Science*, *8579*, 298-311.

Zheng, Y. (2015). Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology*, *6*(3), 29.

2

Visual Exploration of Match Performance Based on Football Movement Data Using the Continuous Triangular Model

Modified from: Zhang, P., Beernaerts, J., Zhang, L., & Van de Weghe, N. (2016). Visual exploration of match performance based on football movement data using the Continuous Triangular Model. *Applied Geography*, *76*, 1-13.

Abstract: The rapid development of information and communication technologies has caused a proliferation of rich and voluminous movement data sources, which thereby necessitates further research on the analysis, modelling and visualisation of moving objects. The human performance analysis in the context of sports, based on sports-oriented movement data using geographical approaches, is an exciting new field. Yet, relatively little attention has been devoted to this topic within the GIScience domain. Therefore, this chapter aims to present a research effort in this fresh field based on the movement data obtained from an entire football match. The research focuses on the exploration of match performance, an important issue in the domain of sports analytics, by utilising the Continuous Triangular Model (CTM). In general, the performance of players and teams is explored from a visualisation perspective according to the CTM diagrams of various motion attributes so that potential suggestions for performance improvements and/or tactical arrangements can be provided. More specifically, the motion attributes comprise

several basic motion attributes and one composite motion attribute. The basic motion attributes include one general motion attribute (speed) that is valid for almost all kinds of moving objects, and two specific motion attributes (ball possession and territorial advantage) being particularly meaningful in football. The composite motion attribute (dominance index) is the combination of the three basic motion attributes. Among the CTM diagrams, some are generated by employing corresponding map algebra operators so as to discover extra information. The results demonstrate that the CTM approach is useful in exploring match performance and discovering important information.

2.1 Introduction

With the rapid development of advanced location-aware technologies such as GPS (global positioning system), RFID (radio frequency identification), Bluetooth, WiFi, and image recognition, data related to moving objects can be acquired more easily than ever before. Nowadays, a large variety of types of movement data are in use or already attracted attention, including floating car data (Civilis et al., 2005), pedestrian walking data (Delafontaine et al., 2012; Xu et al., 2013), bicycle data (Vogel et al., 2014), movement data of animals (Laube et al., 2005; Shamoun-Baranes et al., 2012) and natural phenomena such as hurricanes (Lee et al., 2007), as well as data of sports such as football (Kim et al., 2011), basketball (Chin et al., 2005), volleyball (Chakraborty et al., 2012) and even dancing (Chavoshi et al., 2014). This opens up a new era for sports analytics, where new perspectives may answer questions where traditional methodologies fail. Given the emergence of this new kind of sports data, a number of researchers, particularly those from the fields of computer science, sports analytics and even GIScience (geographical information science), have conducted research from their own perspective. As the focus of this chapter is on football movement data, a review of the typical work that has been done in this area is given.

In the field of computer science, extensive efforts have already been made with regard to the high precision tracking of the positions of both the players and the ball, the extraction of corresponding trajectories from match videos and simple analyses based on the extracted trajectory data. For example, Iwase & Saito (2002) proposed a method of tracking a specific football player by using multiple cameras. Since by means of this method only one player can be tracked, they subsequently extended their work to enable the tracking of multiple players based on homography among multiple views to avoid the occlusion problem of players (Iwase & Saito, 2003). Later, a model that is based on the relationships between the players and the ball was proposed to quantitatively analyse and evaluate the performance of players (Kang et al., 2006). Kim et al. (2011) introduced a framework for the tactic analysis of football matches based on the trajectories of both the players and the ball using spatial and spatio-temporal approaches. As for the work of Niu et al. (2012), they first extracted real-world trajectories from a football video and then analysed the attack patterns that could be found in these trajectories. Lucey et al. (2013) presented an approach to assess team strategy based on an entire season of ball tracking data from the English Premier League (2010 - 2011 season) and investigated possible reasons why the home advantage exists in football. Recently, a collection of spatio-temporal tools specifically for the performance analysis of football players and teams was developed by Gudmundsson & Wolle (2014).

Within the sports analytics field, the first regular occurrences of automated football analysis appeared about a decade ago (Amisco; ProZone; Gudmundsson & Wolle, 2014). These occurrences resulted in the emergence of a large amount of methods and tools aimed at various analysis of football. Among them, the performance of both the players and teams are of great importance to researchers. As such, a football interaction and process model was developed by Beetz et al. (2005) to acquire, interpret and analyse the game using a real-time positioning system. Zhu et al. (2009) proposed a novel approach for extracting tactic information from football video and presented the events in a tactic mode to the coaches and sports professionals. Wisbey et al. (2010) quantified the movement patterns of AFL (Australian Football League) football and the differences among players playing at different positions in order to investigate whether the physical demands had increased over a four season period based on GPS data of the matches. Sampaio & Maçãs (2012) explored how to use football players' dynamic positional data to assess the tactic behaviour by the measurement of movement patterns and inter-player coordination.

In contrast, in spite of the fact that a number of visualisation tools and methods have been developed in the GIScience field towards other types of movement data (Andrienko & Andrienko, 2011; Andrienko & Andrienko, 2012; Andrienko et al., 2012; Wang et al., 2014), these tools and methods are seldom applied in football movement data . Within this field, the work of Laube et al. (2005) takes a central role. In their paper, they presented a generic geographical knowledge discovery approach called REMO (Relative Motion) to explore the motion of moving objects and applied it on a movement dataset consisting of 11 football players covering a time frame of approximately 33 seconds. However, the dataset in their paper is relatively small and they only performed a rather simple analysis, which seems not enough for coaches to grasp a whole match in order to improve corresponding tactics.

Built upon these efforts, the central aim of this chapter is to explore human performance, one of the most important activities in nature, by visualising a number of important motion attributes in football match performance evaluation so that they might better serve coaches and other sports professionals. The adopted tool for visualisation is called the Continuous Triangular Model (CTM), a geographical approach which mainly focuses on time. We aim to apply the CTM to the football analytics domain to extend the applications of the CTM and we believe that this research can provide distinctive added values to the current research in the domain of sports analytics.

The remainder of this chapter is organised as follows. Section 2.2 gives an introduction to the CTM. Section 2.3 introduces the dataset used in this chapter and elucidates the processing of these data. In section 2.4, the match performance is explored in terms of the CTM diagrams generated based on the basic motion attributes. Some of the CTM diagrams are transformed further using related map algebra operators so as to discover extra information. Section 2.5 introduces a composite attribute that is derived by combining multiple basic motion attributes and gives the performance exploration based on the CTM diagrams of the composite motion attribute. In section 2.6, some additional advantages of the CTM in the application domain of sports analytics are discussed. Finally, in section 2.7, the conclusions and recommendations for future work are described.

2.2 The Continuous Triangular Model (CTM)

The Continuous Triangular Model (CTM) is an extension of a 2D representation of time intervals – the Triangular Model (TM) (Qiang et al., 2010; Qiang et al., 2012; Qiang et al., 2012; Qiang et al., 2014), which was originally introduced by Kulpa (Kulpa, 1997; Kulpa, 2006). In the TM, a time interval I (bounded by a start point I^- and an end point I^+) is represented by the intersection point of two straight lines L_1 and L_2, with L_1 passing through I^- and L_2 passing through I^+ (see Figure 2.1). In Figure 2.1, $\alpha_1 = \alpha_2 = \alpha$, with α_1 being the angle between L_1 and the horizontal axis (along the same direction with L_1), α_2 being the angle between L_2 and the horizontal axis (along the same direction with L_2) and α being constant for the entire conceptual space. Thus, the point I in Figure 2.1 is considered as an equivalence of the time interval $[I^-, I^+]$. This way, the horizontal position of the point $mid(I)$ indicates the midpoint of this interval and the vertical position h represents its duration $dur(I)$. In theory, α can be any value in (0°, 90°). However, since the fluctuation of α can cause various space problems on paper, we suggest an angle of 45°. In order to understand the TM with comparative ease, Figure 2.2 gives a supplementary illustration on how to represent time intervals using the TM.

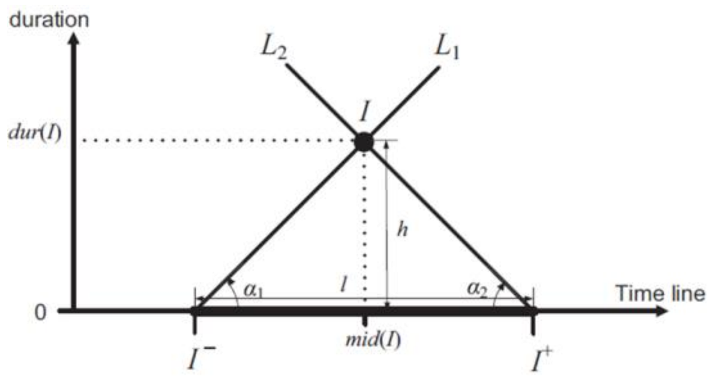
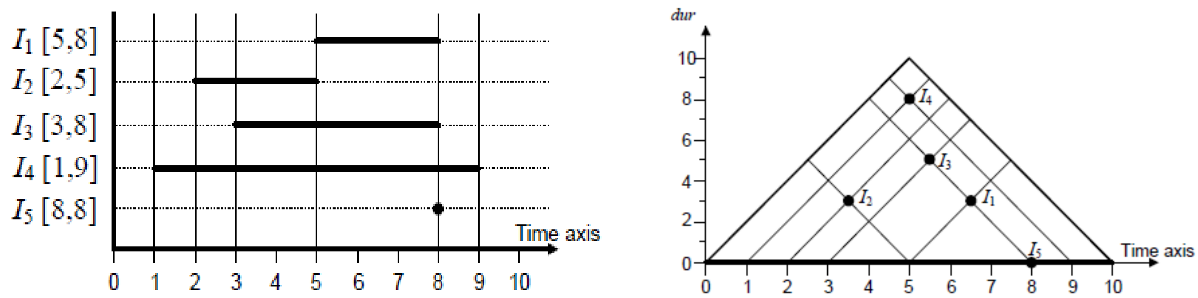


Figure 2.1. Construction of an interval in the Triangular Model (Qiang et al., 2014).



(a) (b)

Figure 2.2. Representation of time intervals using (a) the classical linear model, and (b) the TM (Qiang et al., 2012).

As can be seen in Figure 2.2, the TM provides significant benefits in representing (complex configurations of) time intervals by transforming a line (figure (a)) into a point (figure (b)). This representation results in a collection of discrete points, each of which represents a specific time interval. However, it appears incapable to represent the time intervals continuously. To make this possible, Qiang et al. (2014) extended the TM to the CTM. The core principle of this extension is as follows. Given a time interval *I*, all time intervals during *I* are enclosed in a triangular zone below it, as is shown in Figure 2.3. Thus, every point in this triangular zone corresponds to a specific time interval within the time interval *I*. If assigning to every point a certain value related to the time interval it represents (e.g., $f(I_i)$), the triangular zone can be filled and thus becomes a continuous field. Note that $f(I_i)$ can represent different functions (such as the average, the summation or the standard deviation). Through colour-coding, the continuous field can be displayed as an image, in which every point represents a specific time interval and the colour at that point corresponds to the value of an attribute within this time interval.

Figure 2.4 presents a simple example of representing linear data using the CTM. In Figure 2.4, figure (a) is a linear representation of the speed of a player during five minutes, and figure (b) is the corresponding representation of this linear data using the CTM. For example, the part of curve marked by the dotted red rectangle in figure (a) is represented by the dotted red triangular in figure (b). From figure (b) one can observe that every single point in the CTM has a specific colour that corresponds to a speed value. As each point in the CTM represents a specific time interval, the CTM has the important property that it can simultaneously display the dataset at all temporal resolutions.

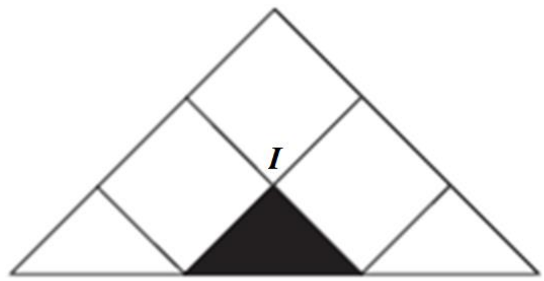
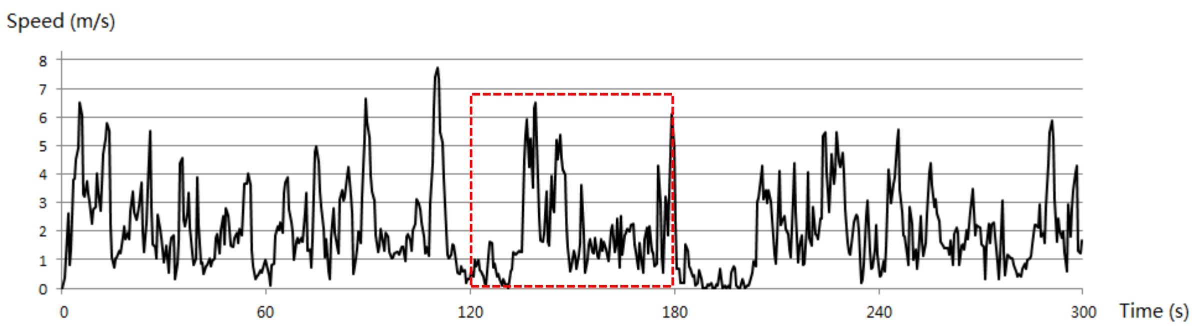
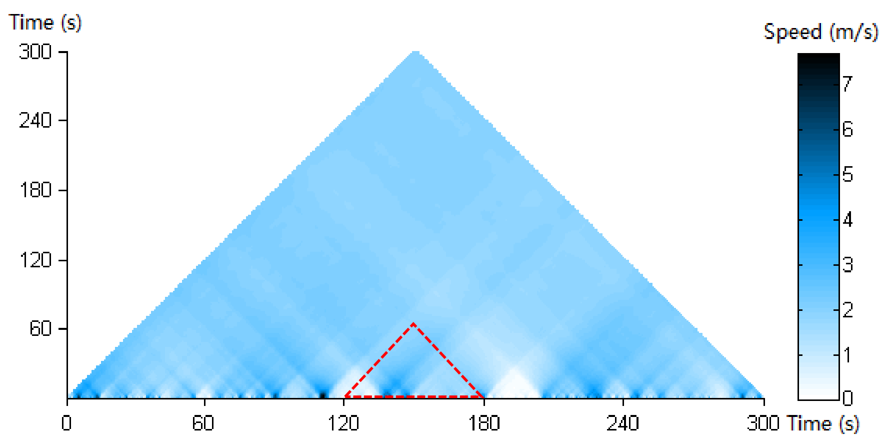


Figure 2.3. Illustration of all time intervals during *I* (represented by the triangle in black).



(a)



(b)

Figure 2.4. Illustration of representing linear data using the CTM: (a) the linear representation of speed data, and (b) the CTM representation of the linear data.

The abilities of the CTM can be enhanced by the support of map algebra operations. Map algebra was first introduced in the late 1970s (Tomlin & Berry, 1979). It is a set of primitive operations in a geographical information system which allows two or more raster layers ('maps') of similar dimensions to produce a new raster layer ('map') using algebra operators. Tomlin (1990) proposed numerous operators, among which the most commonly used ones include addition, subtraction, maximum, minimum, equal to, greater than, and so on. These can be considered as local operators. Besides local operators, focal operators, zonal operators, and global operators are also available. Additionally, different algebra operations can be combined to reveal different kinds of information.

2.3 Dataset and motion attributes

2.3.1 Dataset

The dataset used in this chapter comes from a real football match between 'Club Brugge KV' and 'Standard de Liège' of the Belgian Jupiler Pro League (the top league competition for association football clubs in Belgium), which took place on the second of March 2014. For simplicity, we call them respectively 'Club Brugge' and 'Standard Liège' in the remainder of the chapter. Note that Club Brugge won the match by 1-0 by successfully scoring a goal in around the 77th minute. The two teams ranked respectively first and second at the moment of playing. In this dataset, the (x, y) positions of all players during the whole match were tracked at a temporal resolution of 0.1 seconds. In the original dataset, 1,260,782 (625,131 in the first half and 635,651 in the second half) discrete points with a format of (id, x, y, t) were contained, where *id* identifies a specific player, x and y denote the x and y coordinates of a player's position and t denotes the timestamp when this position was recorded. Besides the spatial and temporal information, the semantic information is included as well, such as the information of both teams and the events that happened during the match.

Given the vast volume of the dataset and according to the needs in this chapter, the dataset is changed to a temporal resolution of 1 second. A small extract of the dataset (after interpolation) is illustrated in Table 2.1. Note that the (x, y) value of the middle point of the football pitch is considered as (0, 0). The structure of the information of both teams is shown in Table 2.2 and some of the important events that happened during the match are listed in Table 2.3.

Table 2.1. An illustration of the dataset.

id	*x* (cm)	*y* (cm)	*t* (s)
342459	-970	-130	0
342459	-950	-120	1
342459	-940	-110	2
…	…	…	…
348909	-990	760	0
348909	-980	760	1
348909	-980	760	2
…	…	…	…
348909	2820	1380	5730
…	…	…	…

Table 2.2. The structure of the information of both teams.

TeamName	IdActor	Occupation	JerseyNumber	Position	IsStarter
Club Brugge	348909	Player	32	Central midfielder right	True
Club Brugge	390565	Player	21	Goalkeeper	True
…	…	…	…	…	…
Standard Liège	343253	Player	37	Wide defender left	True
Standard Liège	379872	Player	39	Centre forward right	True
…	…	…	…	…	…

Table 2.3. An illustration of important events that happened during the match.

EventTime	IdActor1	IdActor2	EventName	BodyPart	LocationX	LocationY
35988	364737	373607	Reception	foot	-2860	-3150
72818	378540		Running with ball	foot	280	3170
173295	373607	379259	Pass	foot	-1320	-230
1993335	348909		Yellow card			
2300226	373607		Off side		-1960	1240
2835701	349993	390565	Shot on target	Header	-4530	340
…	…	…	…	…	…	…

2.3.2 Motion attributes

2.3.2.1 Speed

Speed is a common attribute that is popularly used. The speed during a time interval equals to the value the total distance during the time interval divided by the corresponding time duration.

2.3.2.2 Ball possession

Among the most fundamental aspects of a football game and the performance of players is the passing of the ball (Gudmundsson & Wolle, 2014). Therefore, in this chapter, we define ball possession as the percentage of successful passes made by a team, which is currently popularly adopted by data providers, either TV companies for live games or specialists like Opta, one of the world's leading sports data companies (Opta, 2011). The ball possession of both teams is calculated according to equations (2.1) and (2.2), respectively:

$$p_A = \frac{N_A}{N_A + N_B} \tag{2.1}$$

$$p_B = \frac{N_B}{N_A + N_B} \tag{2.2}$$

where p_A and p_B represent the ball possession of team *A* and team *B*, N_A and N_B represent the number of successful passes made by team *A* and team *B*, respectively.

2.3.2.3 Territorial advantage

Territorial advantage denotes the dominance of a team on the football pitch. It is measured by the distance between the centroid of a team (excluding the goalkeeper) and the edge of the pitch where the scoring goal of the opponent team lies. For example, assume an 11vs11 match in Figure 2.5, where the goalkeepers are represented by open circles and the field players are represented by closed circles in the colour of their team, and the scoring directions of the red and the blue teams are respectively located to the left and to the right. Thus, the length of the red line is considered as the index of the territorial advantage of the red team, and the length of the blue line the blue team. Note that the longer the line is, the stronger the dominance of the team on the field is. Assume the sets of centroids of a team are represented as in half 1 and in half 2, the *x* coordinate values for the edge of the field at the scoring goal direction is X_{edge_1} in half 1 and X_{edge_2} in half 2, the values of territorial advantage at timestamp *i* in half 1 and timestamp *j* in half 2 are calculated based on equations (2.3) and (2.4), respectively:

$$f_i = \left| x_i - \mathrm{X}_{\mathrm{edge\_2}} \right| \tag{2.3}$$

$$f_j = \left| x_j - \mathrm{X}_{\mathrm{edge\_1}} \right| \tag{2.4}$$

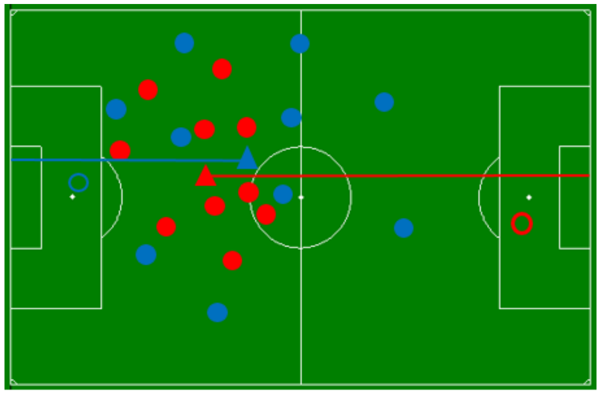


Figure 2.5. An illustration of the territorial advantage: the longer line (in red) indicates that the red team has a stronger territorial advantage, correspondingly the blue team has a weaker territorial advantage.

2.4 Performance exploration based on basic motion attributes

2.4.1 Performance exploration in terms of general motion attribute

2.4.1.1 Speed

Speed is the most common attribute analysed in the existing literature (Wallace & Norton, 2014, Carling et al., 2015). It is meaningful for an individual player, a group of players or a whole team. All these can be visualised using the CTM. In the following, the speed attribute of the goal scorer, the group of midfielders (player b_1, player b_2 and player b_3) and both teams will be explored. For simplicity, the names of players in Club Brugge are replaced by b_1, b_2, b_3 and so on and in Standard Liège' s_1, s_2, s_3 and so on. Besides, the same substitution denotes the same player throughout the whole chapter.

The CTM diagrams of the goal scorer are shown in Figure 2.6. Figure 2.6(a) denotes the average speed of the player at all temporal levels during the whole match. We can observe that fewer fluctuations appear at higher levels (e.g., longer than about 4 minutes) and more fluctuations exist at lower levels (e.g., shorter than about 4 minutes). Therefore, more variations in performance can be investigated within short time intervals. In figure (b), intervals shorter than 4 minutes are considered. In this figure, the fluctuations of speed are explicitly displayed. The dark areas represent that the player ran intensively during the corresponding time intervals, which might mean that he was attacking or defending. The light areas indicate that the player was not quite active during the corresponding time intervals. The details during even shorter time intervals (e.g., during seconds) can be displayed as well. In figure (c), speed during the time intervals no more than 9 seconds is demonstrated. According to figure (c), we can infer that the player performed relatively better in terms of speed at around the 10^{th}, the 14^{th}, the 21^{st}, the 28^{th}, the 62^{nd}, the 67^{th}, and the 80^{th} minute, during a time span no more than 9 seconds. Besides, as the player scored the goal at around the 77^{th} minute, we also visualised the changes in speed from the 75^{th} minute to the 79^{th} minute, in order to investigate his performance during the time intervals before and after the goal. This visualisation is shown in figure (d). Figure (d) clearly exhibits that the average speed from approximately the 76.5^{th} to the 76.7^{th} minute is obviously larger than that during other time intervals. It indicates that the ball was shot into the net and the player was running fast subsequently celebrating his goal during this period. In the following, he experienced a short period with relatively low speed from the 76.7^{th} minute, which indicates the continuing of the match after the temporary celebration.

The speed of the group of midfielders can also be visualised and compared with each other using types of CTM diagrams similar to Figure 2.6(a). However, as one might observe, in Figure 2.6(a), the changes of speed at high temporal levels are not obvious. This makes a visual comparison difficult. Hence, we propose to visualise the values of the CTM diagrams at each temporal level using the same colour-coding, meaning that for each temporal level, the minimum speed value is white, the maximum speed value is black, and the in-between values are coloured by a corresponding colour from light blue to dark blue. Thus, in the CTM diagram, darker colours mean that the player has a higher speed during the corresponding time intervals, while lighter colours represent the opposite. The three CTM diagrams denoting respectively the speed of the three midfielders are generated based on this method. They are shown in figures (a), (b) and (c) in Figure 2.7. From figure (a), we can infer that in general player b_1 ran relatively fast from the beginning to about the 30^{th} minute and from about the 32^{nd} minute to about the 80^{th} minute. Figure (b) shows that

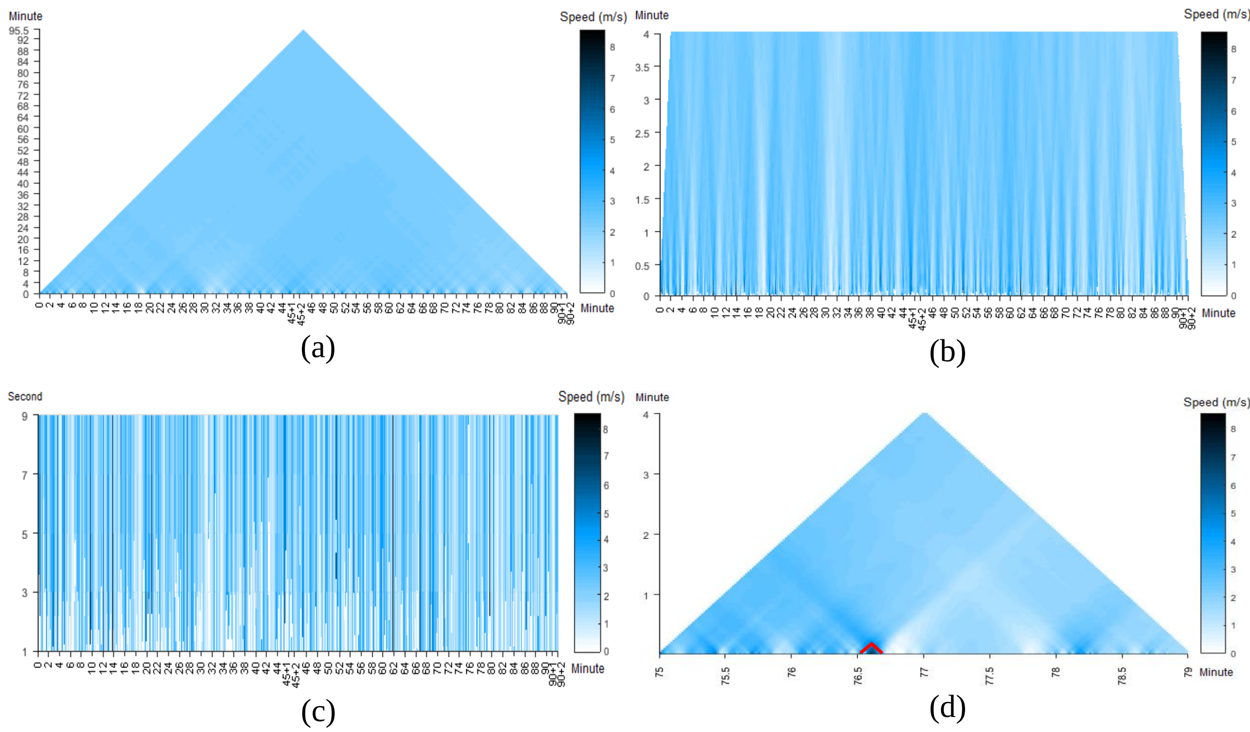


Figure 2.6. The CTM diagrams of the goal scorer during: (a) all time intervals, (b) all time intervals no more than 4 minutes, (c) all time intervals no more than 9 seconds, and (d) all time intervals between the 75th minute and the 79th minute.

player b_2 had a relatively large speed from about the 32nd minute to the end of the match, which indicates that player b_2 performed more active in terms of speed during this period than during the period before. Player b_3 also ran relatively fast from about the 32nd minute to the end, which can be observed in figure (c). Note that there was a pause from about the 30th minute to about the 32nd minute in the match, which makes the speed of the three players similar during this period.

As figures (a), (b) and (c) can only visualise the relative values of speed, they cannot be used to precisely compare which player ran faster than the other, and during which time intervals. However, with map algebra operators, this can be achieved. In this case, the operator of subtraction is adopted. Based on the subtraction operator, three new CTM diagrams are generated, as shown in figures (d), (e) and (f). Figure (d) shows that player b_1 had an advantage in performance in terms of speed during long time intervals (e.g., longer than about 4 minutes) when compared to player b_2, but during short time intervals

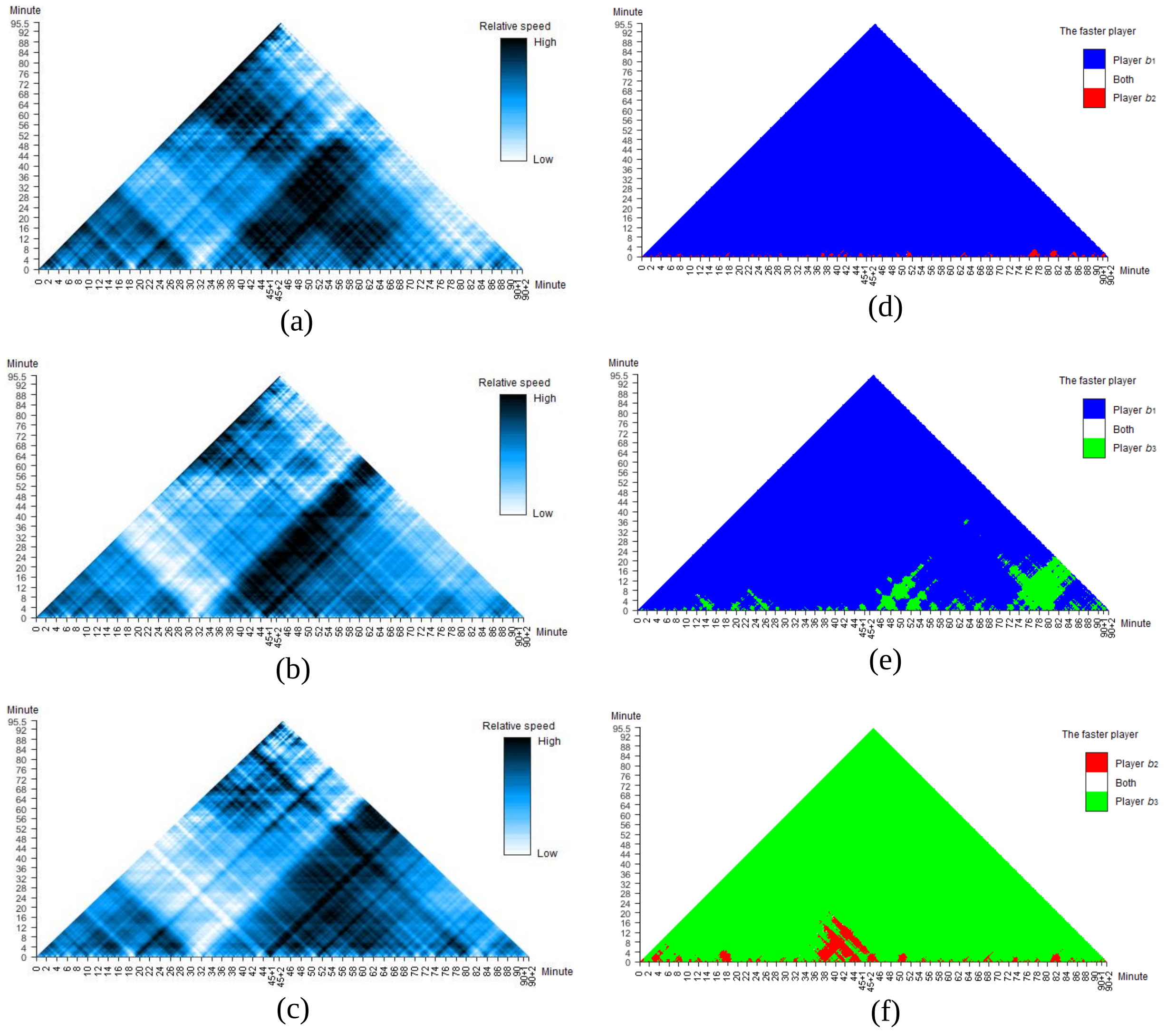


Figure 2.7. The CTM diagrams of the three midfielders: (a) player b_1, (b) player b_2, (c) player b_3, (d) player b_1 compared with player b_2, (e) player b_1 compared with player b_3, and (f) player b_2 compared with player b_3.

(e.g., less than about 4 minutes), each player had his own dominant periods during which he performed better in speed. Based on this, we can infer that player b_2 might be better at executing short-term runs than long-term runs. Figure (e) shows that player b_1 also ran faster compared to player b_3 during long time intervals, especially for intervals longer than about 20 minutes. For time intervals shorter than 20 minutes, player b_1 ran faster in the first half while in the second half player b_3 ran faster. Based on figure (f), we can infer that player b_3 ran faster than player b_2 during time intervals longer than about 20 minutes. For time intervals shorter than 20 minutes, player b_2 performed better in speed than the others between the 32nd minute and the end of the first half, even though each player had periods during which he had advantage in speed.

In addition, the overall performance of the three midfielders can be compared together by generating one single CTM diagram using the relational operator of maximum. The generated CTM diagram is shown in Figure 2.8. Based on Figure 2.8, in general, we can conclude that player b_1 performed best in speed during time intervals longer than about 20 minutes among the three players, but during time intervals shorter than 20 minutes, player b_1 performed better in speed in the first half while player b_3 better in speed in the second half. However, player b_2 had his own advantage in performance according to speed during quite short time intervals (e.g., shorter than about 4 minutes).

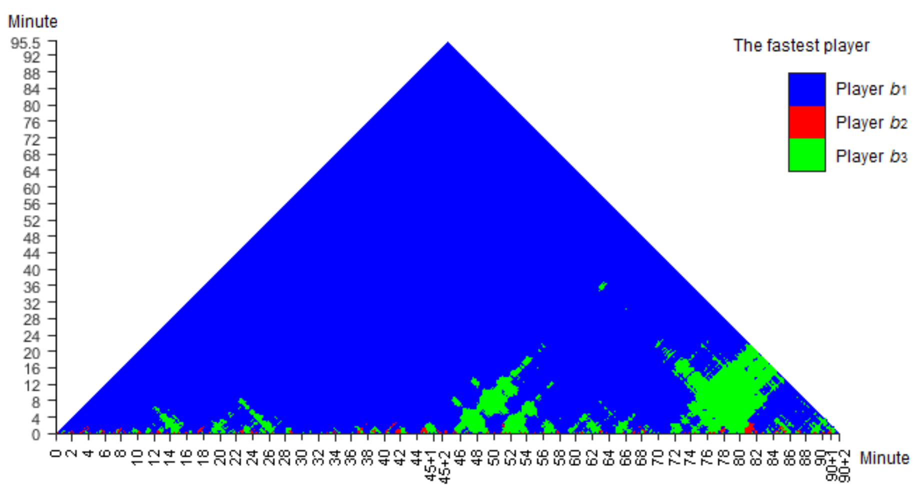


Figure 2.8. The CTM diagram of the comparison of the three midfielders of player b_1, player b_2 and player b_3.

The CTM diagram denoting the speed performance of both teams is generated with the subtraction operator and displayed in Figure 2.9. Figure 2.9 shows that Club Brugge was dominant in speed during time intervals longer than about 24 minutes, and Standard Liège had its advantage in speed during time intervals shorter than 24 minutes. From the legend of Figure 2.9, we can notice that the speed of Club Brugge was about 3.6 m/s higher than that of Standard Liège at most during the same time interval, but Standard Liège was about

2.0 m/s higher than that of Club Brugge at most. During time intervals shorter than 24 minutes, the overall performance of Standard Liège in speed was obviously better during the periods from the beginning to the 14th minute, from the 22nd minute to the 36th minute, from the 38th minute to the 50th minute, from the 54th minute to the 60th minute, from the 70th minute to the 74th minute, from the 78th minute to the 82nd minute and from the 83rd minute to the 86th minute. In addition, the colour from the 75th minute to the 78th minute is relatively dark, which means that Club Brugge was relatively more dominant in speed during this period. This is probably due to that the players of Club Brugge ran fast trying to score, and after scoring, they all ran fast towards the goal scorer for celebration.

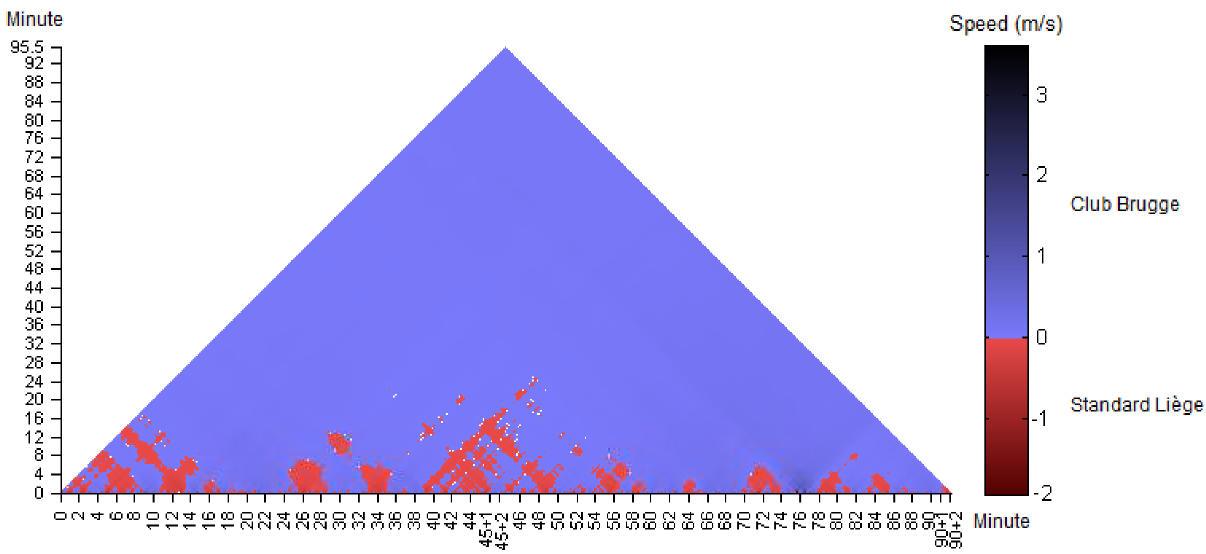
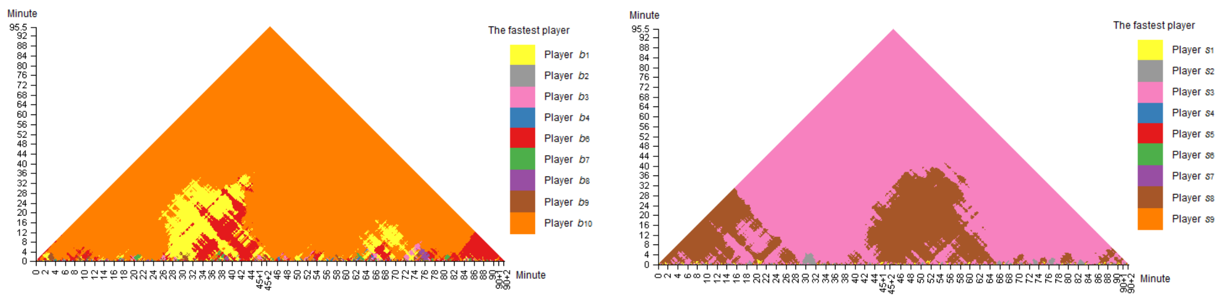


Figure 2.9. The CTM diagram of the speed of both teams after subtraction operation (the speed of Club Brugge subtracts that of Standard Liège): blue colour means that the speed of Club Brugge is higher and red colour means that the speed of Standard Liège is higher (the darker the colour is, the higher the speed is), a positive value (e.g., *n*) denotes that the speed of Club Brugge is *n* m/s faster than that of Standard Liège, a negative value (e.g., -*n*) denotes that the speed of Club Brugge is *n* m/s slower than that of Standard Liège, and 0 means that both teams have the same speed

The performance of players in each team can also be explored by generating the CTM diagrams according to their speed by means of the relational operator of maximum to determine which player has the largest speed and during each time interval. The fastest players during each time interval at all temporal scales for both Club Brugge and Standard Liège are shown in Figure 2.10. In Figure 2.10, figure (a) corresponds to the fastest players in Club Brugge during all time intervals and figure (b) in Standard Liège. From Figure 2.10 we can observe that players b_{10}, b_1 and b_6 were the three fastest players on a whole in Club Brugge while in Standard Liège the two players who had obvious advantages in speed were player s_3 and player s_8. This may indicate that these players might play a relatively important role in their corresponding teams and thus deserved specific attention during this match.



(a) (b)

Figure 2.10. The fastest players in both teams: (a) Club Brugge, and (b) Standard Liège.

2.4.2 Performance exploration in terms of specific motion attributes

2.4.2.1 Ball possession

The ball possession of a team denotes the ratio between the successful passes executed by this team and the sum of successful passes executed by both teams. The CTM diagram of the comparison of both teams on ball possession is shown in Figure 2.11. It can be observed that, on a whole, Standard Liège had more ball possession than Club Brugge during the whole match, which implies that Standard Liège had more passes in this match. Specifically, during the first half, Standard Liège had an obvious advantage in ball possession during the time intervals from the 2nd minute to the 12nd minute and from the 36th minute to the end of the first half on a whole, while Club Brugge dominated ball possession during the period from the 12th minute to the 28th minute. During the second half, Club Brugge performed better in terms of ball possession from the beginning to the 68th minute and from the 70th minute to the 83rd minute, while Standard Liège performed better from the 68th minute to the 70th minute and from the 83rd minute to the end. Besides, Standard Liège appeared to perform better in ball possession in relatively long time intervals while Club Brugge better in relatively short time intervals.

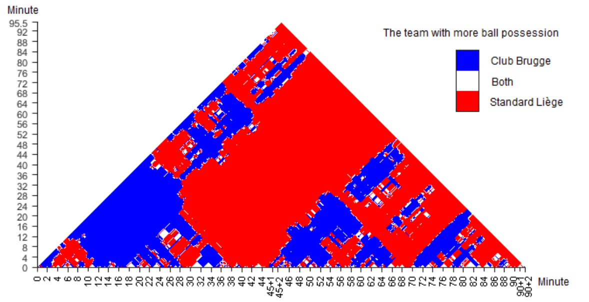


Figure 2.11. The CTM diagram of the comparison of ball possession for both teams.

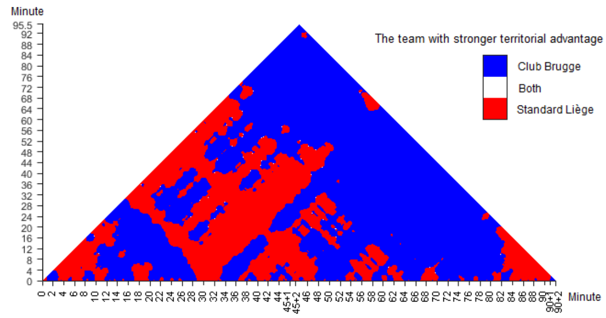


Figure 2.12. The CTM diagram of the comparison of territorial advantage for both teams.

2.4.2.2 Territorial advantage

The territorial advantage denotes the dominance of a team on the football pitch, and it can be considered as a measure of the aggression abilities of a team. When the territorial advantage is larger, the team tends to be attacking on a whole. However, similar to ball possession, a stronger territorial advantage cannot always mean a final win. The CTM diagram of the comparison of territorial advantage for both teams is shown in Figure 2.12. From Figure 2.12 we can see that generally Club Brugge performed better in field dominance than Standard Liège during the whole match, but Standard Liège performed better during the first half while Club Brugge performed better during the second half, although the territorial advantage of both teams oscillated in short time intervals. Besides, Club Brugge appeared to perform better during relatively long time intervals (e.g., longer than about 74 minutes). However, during the time intervals shorter than 74 minutes, Standard Liège also had its significant dominant periods in field, such as from the 0th minute to the 20th minute, from the 27th minute to the 33rd minute, from the 36th minute to the 50th minute, from the 52nd minute to the 64th minute, from the 66th minute to the 71st minute, and from the 83rd minute to the end of the match. Based on the CTM diagram, coaches can try to find out the reasons why the team had good/bad territorial advantage during specific time intervals so that they can arrange corresponding tactics in the future matches.

2.5 Performance exploration based on the composite motion attribute

Composite motion attributes that build upon different types of individual motion attributes can also be employed to serve the performance exploration using the CTM on the basis of corresponding map algebra operations. In this section, the overall dominance of both teams is explored based on a composite motion attribute – dominance index. Note that this index is proposed by us since we haven't found similar index in literature. We consider the dominance index of a team consisting of three variables: ball possession of the team, territorial advantage of the team and average speed of the team, and each of the variables has a corresponding weight. Hence, the dominance index can be represented as equation (2.5):

$$I = w_1 * p + w_2 * f + w_3 * v \tag{2.5}$$

where I, p, f and v respectively denote the dominance index, ball possession, territorial advantage and average speed of the team, and w_1, w_2, w_3 denote the corresponding weights. Note that w_1, w_2 and w_3 are larger than 0 and $w_1 + w_2 + w_3 = 1$.

As three CTM diagrams can be generated based on p, f and v, respectively, a corresponding CTM diagram of I can be derived based on the three CTM diagrams using the map algebra of addition, according to which the overall dominance of the corresponding team can be explored. However, as the units of the three variables are distinct, a normalisation operation should be applied to each variable before executing the map algebra operations so that the three variables have the same range of values. In this chapter, the values are all transformed to a range between 0 and 1.

Figure 2.13 illustrates the specific procedure of generating the CTM diagram of dominance index for Club Brugge. Similarly, the CTM diagram of dominance index for Standard Liège can also be generated. As different weights for the three variables can cause different results, we also investigated the effect under different parameter combinations and found that the results are significantly similar for relatively long time intervals (e.g., longer than about 5 minutes), although some differences exist for short time intervals. This indicates that the dominance index is robust in weights setting, thus is suitable for exploring the overall dominance abilities, especially from the perspective of long time intervals.

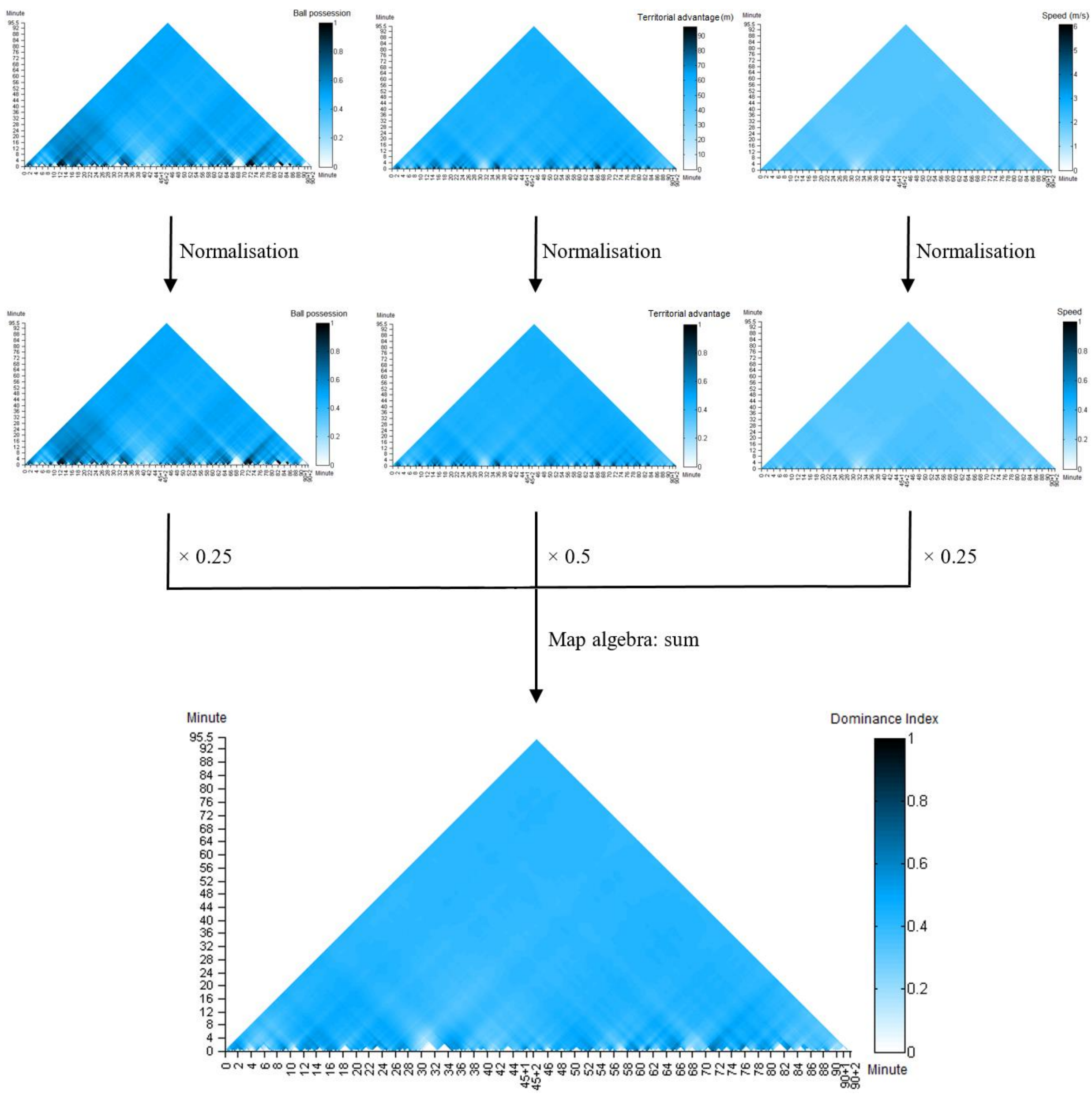


Figure 2.13. The procedure of generating the CTM diagram of dominance index for Club Brugge.

In this chapter, we take the values of the three weights as an example: $w_1 = 0.25$, $w_2 = 0.5$ and $w_3 = 0.25$. The comparison of both teams on the performance of overall dominance can be investigated according to the CTM diagram displayed in Figure 2.14, which is generated based on the two corresponding CTM diagrams of dominance index by employing the map algebra operator of subtraction. According to Figure 2.14 we can conclude that Club Brugge had a stronger overall dominance ability than Standard Liège for relatively long (e.g., longer than about 24 minutes) periods. During the time intervals between 2 minutes and 24 minutes, both teams had their own time intervals during which the overall dominance ability of the team was stronger. However, it seems that the overall dominance ability of Club Brugge was stronger than that of Standard Liège on a whole. During the time intervals less than 2 minutes, it appears that during many time intervals (i.e., the time intervals in white), the overall dominance ability of Club Brugge was equivalent to that of Standard Liège. However, during the remaining time intervals (i.e., the time intervals in blue or red), the overall dominance of Club Brugge is stronger than that of Standard Liège. Hence, in all, during the time intervals less than 2 minutes, the overall dominance of Club Brugge is stronger. However, as the time intervals became longer, both teams had their dominant time intervals, but relatively Club Brugge had stronger overall dominance abilities than Standard Liège. Besides, when the time intervals became even much longer, Club Brugge had overwhelming overall dominance ability compared to Standard Liège. This indicates that Club Brugge revealed stronger overall dominance ability than Standard Liège.

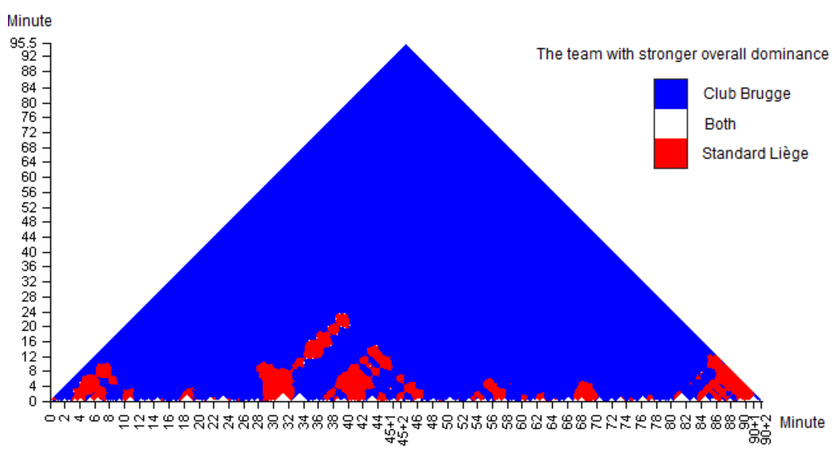


Figure 2.14. The CTM diagram of dominance index for both teams after subtraction operation.

2.6 Discussion

As a key contribution, this chapter addresses the application of the Continuous Triangular Model into the domain of sports analytics to explore match performance. The results show that the CTM can provide distinctive information about match performance to sports professionals and it has the potential to be used in sports analytics to support tactics decision making. Besides, the CTM can also reveal its advantage in other ways as follows.

The CTM can be used to solve current practical problems. Currently, to better analyse the match and to acquire more information about the movement and performance of players, clubs usually first divide a match video into several short episodes (normally 15 minutes per episode, such as from the 0^{th} minute to the 15^{th} minute, from the 15^{th} minute to the 30^{th} minute) and then analyse the match episode by episode. In this way, insights can be gained into the overall movement and performance of the players during each episode. This method, however, fails to automatically generate insights, which are related to time intervals that do not map directly to this predefined set of episodes (e.g., an attack which starts in one episode and ends in the next). Besides, the events that occur over the boundaries of an episode cannot be recognised, since the event is divided into subparts. The CTM approach, in contrast, does not necessitate such predefined sets of episodes, since the CTM diagrams can visualise the movement during any time interval at any temporal scale. Take Figure 2.15 for example to compare both methods. With the method currently

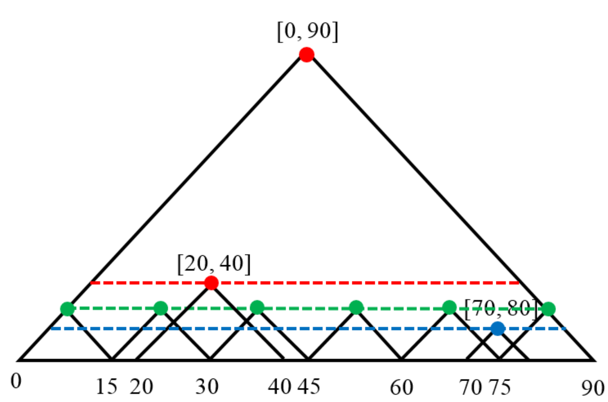


Figure 2.15. The comparison between the CTM approach and currently used method.

adopted by the clubs, typically the overall movement of the players over the six intervals (represented with green closed circles) are explored. However, with the CTM approach, the movement during all time intervals on the temporal scales below the green dotted line (e.g., the blue dotted line), above it (e.g., the red dotted line) and in between can be explored.

Validation of the effectiveness of the results is another important task. Since the world wide popularity of football, lots of related reports can be found either on websites, in newspapers or magazines, especially for the matches of famous teams. Among the reports, many are textual descriptions. We can confront these textual descriptions to check whether the information/knowledge that we get from the CTM diagrams matches the true cases. For example, we checked the game reports[1] at the official website of Club Brugge and we found two comments 'Les Rouches[2] were the dominant side in the opening minutes' and 'The Blues[3] restored the balance after ten minutes and started dictating the rhythm of the game'. This denotes that the overall dominance abilities of Standard Liège probably appeared stronger than Club Brugge in the first ten minutes and then Club Brugge had stronger overall dominance abilities lasting for several minutes. From Figure 2.14 we can see that the results shown by the CTM diagram coincide well with the reports. Although this example is simple, it shows that the CTM diagrams indeed can be validated with textual descriptions. Thus, people may have a better understanding of the match if combining both the CTM diagrams and the necessary textual descriptions.

The CTM can also be employed to explore semantic information related to movement data. Generally, lots of events can happen during a whole match, such as 'pass', 'reception', 'shot on target', 'shot not on target' and 'goal'. Coaches may be interested in some specific event(s) and wish to know the overall distribution of the event(s) in the whole match at all temporal scales. This can be achieved using the CTM. Take Club Brugge for example, the CTM diagram of the events of 'shot on target' and 'goal' is displayed in Figure 2.16, since

[1] http://clubbrugge.be/en/news/game-reports/17723/club-brugge-defeat-standard-1-0.
[2] Standard Liège.
[3] Club Brugge.

the two events can be considered as one of the important factors whether a team can win or not. Figure 2.16 clearly demonstrates the overall distribution of the number of the two events during any time interval in the match, based on which coaches can analyse related things. Besides, the events between the two teams can also be compared using the CTM. For example, the CTM diagram of the two events ‘shot on target’ and ‘goal’ between Club Brugge and Standard Liège is shown in Figure 2.17, from which we can easily know that Standard Liège had more superiority in the two events on a whole, but Club Brugge was better during the late periods in the second half in the match. Similarly, other basic events can also be investigated using the CTM. This might be an interesting added value to news media (e.g., newspapers) since it is not difficult for ordinary readers to understand.

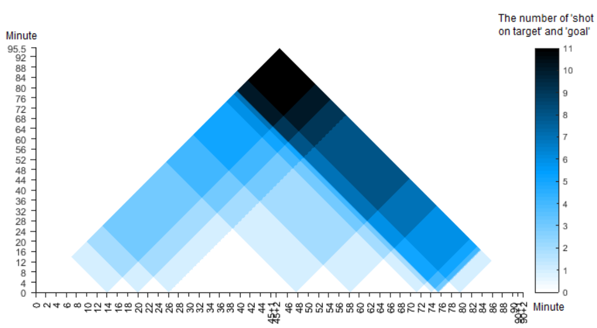


Figure 2.16. The CTM diagram of ‘shot on target’ and ‘goal’ for Club Brugge.

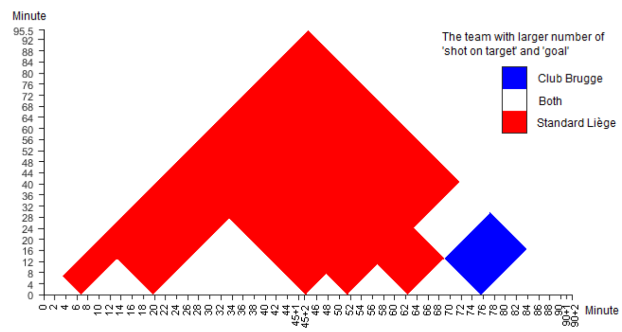


Figure 2.17. The CTM diagram of the comparison of ‘shot on target’ and ‘goal’ for both teams.

2.7 Conclusions and future work

In this chapter, we propose to employ a rather novel approach, the Continuous Triangular Model, to explore match performance based on the movement data obtained from a complete football match in order to provide some currently under investigated values to sports professionals (especially coaches) from a GIScience perspective. This is strongly stimulated by the fast development of advanced location-aware technologies in recent years, which makes it easier to obtain movement data of a complete sports match. We aim to extend the application of geographical approaches to the domain of sports analytics to solve domain problems. Generally speaking, the work involved in this chapter focuses on a distinctive characteristic of the CTM – visualisation. Specifically, we explore the performance of both players and teams by means of visualising their various motion attributes. As for the motion attributes, we visualise not only the basic attributes, but also a more complex composite attribute which builds upon multiple basic ones. Some further operations are also executed to the basic motion attributes based on map algebra operators, by which some useful knowledge is discovered. The results show that the CTM approach from the GIScience domain can indeed be used to serve the domain of sports analytics, especially the match performance exploration, by which sports professional can get some distinctive information that traditional approaches in sports analytics cannot or are not easy to find. According to the distinctive information discovered by the CTM and other information discovered using traditional sports analytics approaches, coaches can improve tactics further and arrange tactics more flexible than ever before in the future in order to try all the best to win the matches. For example, according to Figure 2.10(a), player b_{10} can be considered as a potential (best) player if the coach wishes to choose a player with fast running to execute this tactic. We believe that this new approach can be used to better serve the domain of sports analytics in the future.

Due to the strong extensibility and the various functions the CTM has, we summarise a number of avenues for future research. One avenue is the incorporation of the methodology in an interactive graphical user interface. This will greatly aid in augmenting the usability of the approach for sports analytics experts. Another avenue is that the CTM can be incorporated with a heat map, a useful and popular visualisation tool in football analysis, so as to enhance the abilities of the basic heat map. Thus, the weak ability of heat map in coping with temporal information can be made up for.

References

Amisco. Available from: http://www.amisco.eu.

Andrienko, G., Andrienko, N., Burch, M., & Weiskopf, D. (2012). Visual analytics methodology for eye movement studies. *IEEE Transactions on Visualization and*

Computer Graphics, *18*(12), 2889-2898.

Andrienko, N., & Andrienko, G. (2011). Spatial generalization and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics*, *17*(2), 205-219.

Andrienko, N., & Andrienko, G. (2012). Visual analytics of movement: an overview of methods, tools and procedures. *Information Visualization*, *12*(1), 3-24.

Beetz, M., Kirchlechner, B., & Lames, M. (2005). Computerized real-time analysis of football games. *IEEE Pervasive Computing*, *43*(3), 33-39.

Carling, C., Gregson, W., McCall, A., Moreira, A., Wong, D. P., & Bradley, P. S. (2015). Match running performance during fixture congestion in elite soccer: research issues and future directions. *Sports Medicine*, *45*(5), 605-613.

Chakraborty, B., & Meher, S. (2012). A trajectory-based ball detection and tracking system with applications to shot-type identification in volleyball videos. In *Proceedings of the 2012 International Conference on Signal Processing and Communications* (pp. 1-5).

Chavoshi, S. H., De Baets, B., Neutens, T., Ban, H., Ahlqvist, O., De Tré, G., & Van de Weghe, N. (2014). Knowledge discovery in choreographic data using Relative Motion matrices and Dynamic Time Warping. *Applied Geography*, *47*, 111-124.

Chin, S. L., Huang, C. H., Tang, C. T., & Hung, J. C. (2005). An application based on spatial-relationship to basketball defensive strategies. *Lecture Notes in Computer Science*, *3823*, 180-188.

Civilis, A., Jensen, C. S., & Pakalnis, S. (2005). Techniques for efficient road-network-based tracking of moving objects. *IEEE Transactions on Knowledge and Data Engineering*, *17*(5), 698-712.

Delafontaine, M., Versichele, M., Neutens, T., & Van de Weghe, N. (2012). Analysing spatiotemporal sequences in Bluetooth tracking data. *Applied Geography*, *34*, 659-668.

Gudmundsson, J., & Wolle, T. (2014). Football analysis using spatio-temporal tools. *Computers, Environment and Urban Systems*, *47*, 16-27.

Iwase, S., & Saito, H. (2002). Tracking soccer player using multiple views. In *MVA* (pp. 102-105).

Iwase, S., & Saito, H. (2003). Tracking soccer players based on homography among multiple views. In *Proceedings of Visual Communications and Image* (pp. 283-292).

Kang, C., Hwang, J., & Li, K. (2006). Trajectory analysis for soccer players. In *Proceedings of the 6th IEEE International Conference on Data Mining* (pp. 377-381).

Kim, H., Kwon, O., & Li, K. (2011). Spatial and spatiotemporal analysis of soccer. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 385-388).

Kulpa, Z. (1997). Diagrammatic representation of interval space in proving theorems about interval relations. *Reliable Computing*, *3*(3), 209-217.

Kulpa, Z. (2006). A diagrammatic approach to investigate interval relations. *Journal of Visual Languages and Computing*, *17*(5), 466-502.

Laube, P., Imfeld, S., & Weibel, R. (2005). Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science*, *19*(6), 639-668.

Lee, J. G., Han, J., & Whang, K. (2007). Trajectory clustering, a partition-and-group framework. In *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data* (pp. 593-604).

Lucey, P., Oliver, D., Carr, P., Roth, J., & Matthews, I. (2013). Assessing team strategy using spatiotemporal data. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1366-1374).

Niu, Z., Gao, X., & Tian, Q. (2012). Tactic analysis based on real-world ball trajectory in soccer video. *Pattern Recognition*, *45*(5), 1937-1947.

Opta. Blog a ball possessed. Retrieved January 10, 2011, from http://www.optasports.com/news-area/blog-a-ball-possessed.aspx.

ProZone. Available from: http://www.prozonesports.com.

Qiang, Y., Chavoshi, S. H., Logghe, S., De Maeyer, P., & Van de Weghe, N. (2014). Multi-scale analysis of linear data in a two-dimensional space. *Information Visualization*, *13*(3), 248-265.

Qiang, Y., Delafontaine, M., Asmussen, K., Stichelbaut, B., De Tré, G., De Maeyer, P., & Van de Weghe, N. (2010). Modelling imperfect time intervals in a two-dimensional space. *Control and Cybernetics*, *39*(4), 983–1010.

Qiang, Y., Delafontaine, M., Neutens, T., Stichelbaut, B., De Tré, G., De Maeyer, P., & Van de Weghe, N. (2012). Analysing imperfect temporal information in GIS using the triangular model. *The Cartographic Journal*, *49*(3), 265-280.

Qiang, Y., Delafontaine, M., Versichele, M., De Maeyer, P., & Van de Weghe, N. (2012). Interactive analysis of time intervals in a two-dimensional space. *Information Visualization*, *11*(4), 255-272.

Sampaio, J., & Maçãs, V. (2012). Measuring tactical behaviour in football. *International Journal of Sports Medicine*, *33*(5), 395-401.

Shamoun-Baranes, J., van Loon, E. E., Purves, R. S., Speckmann, B., Weiskopf, D., & Camphuysen, C. J. (2012). Analysis and visualization of animal movement. *Biology Letters*, *8*(1), 6-9.

Tomlin, C. D. (1990). *Geographic information systems and cartographic modelling*. Prentice Hall.

Tomlin, C. D., & Berry, J. K. (1979). A mathematical structure for cartographic modeling in environmental analysis. In *Proceedings of the 39th Symposium of the American Congress on Surveying and Mapping*, (pp. 269-283).

Vogel, M., Hamon, R., Lozenguez, G., Merchez, L., Abry, P., Barnier, J., ... & Robardet, C. (2014). From bicycle sharing system movements to users: a typology of Vélo' cyclists in Lyon based on large-scale behavioural dataset. *Journal of Transport Geography*, *41*, 280-291.

Wallace, J., & Norton, K. (2014). Evolution of world cup soccer final games 1966-2010: game structure, speed and play patterns. *Journal of Science and Medicine in Sport*, *17*(2), 223-228.

Wang, Z., Ye, T., Lu, M., Yuan, X., Qu, H., Yuan, J., & Wu, Q. (2014). Visual exploration of sparse traffic trajectory data. *IEEE Transactions on Visualization and Computer*

Graphics, *20*(12), 1813-1822.

Wisbey, B., Montgomery, P. G., Pyne, D. B., & Rattray, B. (2010). Quantifying movement demands of AFL football using GPS tracking. *Journal of Science and Medicine in Sport*, *13*(5), 531-536.

Xu, Z., Sandrasegaran, K., Kong, X., Zhu, X., Zhao, J., Hu, B., & Lin, C. C. (2013). Pedestrian monitoring system using Wi-Fi technology and RSSI based localization. *International Journal of Wireless & Mobile Networks*, *5*(4), 17-34.

Zhu, G., Xu, C., Huang, Q., Rui, Y., Jiang, S., Gao, W., & Yao, H. (2009). Event tactic analysis based on broadcast sport video. *IEEE Transactions on Multimedia*, *11*(1), 49-67.

3

Knowledge Discovery in Movement Data: A Cross-Scale Oriented Sequence Analysis Approach

Modified from: Zhang, P., & Van de Weghe, N. (2018). Knowledge discovery in movement data: a cross-scale oriented sequence analysis approach. *Applied Geography*. (Under review)

Abstract: Recent advances in location-aware technologies have resulted in the ubiquity of movement data. Hence, numerous approaches have been developed within GIScience to explore movement data in order that valuable information can be revealed. Yet, relatively few have considered the cross-scale issue on the temporal dimension. Therefore, this chapter proposes a cross-scale oriented sequence analysis approach for discovering knowledge in movement data by taking temporal scale into account. The key of our approach is the construction of sequences based on the Continuous Triangular Model (CTM), a tool to represent linear data at multiple temporal scales. Two distinct research aims are subsequently derived: investigating the changes of motion attributes of moving objects across different temporal scales and detecting the time intervals during which active events might have occurred. The movement data obtained from an entire football match are employed to validate the effectiveness of the proposed approach. The findings demonstrate that the proposed approach is useful and efficient in discovering knowledge in movement data.

3.1 Introduction

With the recent advances in location-aware technologies, such as GNSS (global navigation satellite system), RFID (radio-frequency identification), Bluetooth sensors, WiFi, and image recognition, the changes of positions of various moving objects over time are becoming more convenient to be tracked than ever before. This directly results in the ubiquity of movement data, such as transportation related movement data (Civilis et al., 2005; Xu et al., 2013; Vogel et al., 2014), animal movement data (Laube et al., 2005; Shamoun-Baranes et al., 2012; Demšar et al., 2015), sports movement data (Gudmundsson & Wolle, 2014; Gomez et al., 2014; Zhang et al., 2016; Zhang et al., 2018), eye movement data (Andrienko et al., 2012), and even natural phenomena movement data (Lee et al., 2007). Due to the sizable amount of movement data in this data-rich era, data mining methods to discover valuable knowledge that can be used for various purposes are large in demand.

Data mining has already attracted a great deal of attention both in the information industry and in society as a whole for a long time, due to the wide availability of huge amounts of data and the urgent needs for turning such data into useful knowledge and information (Han & Kamber, 2006). Specific to movement data mining, we broadly divide the existing methods into three categories: (1) shape-based methods; (2) attribute-based methods, and (3) shape-and-attribute based methods. Shape-based methods essentially focus on the geometric characteristics (i.e., shape) of the trajectory of a movement, as a trajectory can be considered as a series of discrete points in chronological order. Shape-based methods are typically used for trajectory clustering (Lee et al., 2007; Zhang et al., 2014), pattern mining (Laube et al., 2005; Gudmundsson et al., 2007; Andersson et al., 2008; Jeung et al., 2008; Wachowicz et al., 2011; Turdukulov et al., 2014) and outlier detection (Lee et al., 2008; Liu et al., 2012). Attribute-based methods mainly focus on analysing the changes of motion attributes (such as speed, acceleration, distance and direction) which are used to characterise the motions of moving objects over time. They can be used to mine useful patterns (Laube et al., 2005), explore the similarities of trajectories (Dodge et al., 2012;

Chavoshi et al., 2015), and even predict the positions of moving objects over time (Elsner & Kara, 1999). Shape-and-attribute based methods can be considered as a mixture of shape-based methods and attribute-based methods. An important strength of this type of methods is that it can enhance and refine the meanings of trajectories by integrating semantic information (Buchin et al., 2012; Elragal & EL-Gendy, 2013; Buchin et al., 2014).

Among the current methods, those that focus on discovering knowledge at a single spatial or a single temporal scale take a large proportion. As is known, scale is a common problematic issue in many disciplines, particularly those that study phenomena embedded in space and time, e.g., GIScience. In GIScience, scale is of important significance, and mostly denotes resolution or extent (Goodchild, 2011). Geographical data normally have a specific resolution, so the operations (such as transformations and analyses) of such data are scale-specific, since things might vary dramatically as scale changes. Given the importance of scale in geographical data, it even has been considered as the fifth dimension in 5D data modelling (van Oosterom & Stoter, 2010). Thus, it is crucial to take scale into account when analysing space and/or time related data. In movement data, limited research has been undertaken from a cross-scale perspective. Laube & Purves (2011) investigated the changes of three motion attributes (i.e., speed, turning angle and sinuosity) of ten cows at six keenly selected temporal scales. Postlethwaite et al. (2013) presented a new multi-scale measure MSSI (Multi-Scale Straightness Index) to analyse animal movement data at multiple temporal scales. Soleymani et al. (2014) proposed a methodology to explore the behavioural movement of zebrafish by joint spatio-temporal cross-scale analysis of three motion attributes (i.e., speed, acceleration and sinuosity). A prevalent flaw of the current research is that only a very small number of temporal scales have been carefully selected and taken into consideration, which may result in apparent discrepancies between the findings and the real circumstances. More precise findings are capable to be gained with more temporal scales.

Hence, one of the aims in this chapter is to investigate the changes of motion attributes by taking into account many typical temporal scales. Additionally, the changes of motion attributes transformed with a descriptive statistics measure (i.e., mean) across different temporal scales are explored so that more abundant information can be gained. In order to achieve this, a cross-scale oriented sequence analysis approach is proposed. Note that our aim is primarily methodological, which means that we aim not to investigate the detailed changes of all possible motion attributes, but rather demonstrate how the proposed approach can be applied to analyse the changes in motion attributes across different temporal scales in a large volume of movement data. Based on the proposed approach, the other aim of this chapter is derived: investigating the time intervals during which active events might have occurred. This has been rarely conducted in previous research. One similar research, to some extent, is the one presented in (Teimouri et al., 2016), where the authors detected the time intervals during which inactive events might have occurred. However, in this work, the scale issues in the temporal dimension were not considered. Sequence analysis is a promising approach to the analysis of processes, events and changes (Abbott 1990). It has been proverbially used in various application fields, such as pattern recognition (Jain et al., 2000), speech recognition (Sakoe & Chiba, 1978), handwriting recognition (Tappert et al., 1990), as well as in the GIScience domain (Çöltekin et al., 2010; Delafontaine et al., 2012; Yuan & Raubal, 2014). In our approach, three main steps are included. First, the values of the motion attributes at the finest temporal scale are calculated based on the movement data. Second, four types of sequences are constructed based on the Continuous Triangular Model (CTM), initially proposed by Qiang et al. (2014). Third, the discovery of knowledge (i.e., the changes in motion attributes across different temporal scales and the time intervals during which active events might have occurred) is executed based on these four types of sequences. The proposed approach can be classified as an attribute-based movement data mining method, and mainly focuses on the temporal scale.

The remainder of this chapter is organised as follows. In section 3.2, the CTM, on the basis of which the four types of sequences are constructed, is introduced. In section 3.3, the four types of sequences are described in detail. Section 3.4 mainly depicts the knowledge discovery methods. In section 3.5, a case study is conducted using the proposed approach.

The conclusions and future work are summarised in section 3.6.

3.2 The Continuous Triangular Model (CTM)

The Continuous Triangular Model (CTM) is an extension of a 2D representation of time intervals, the Triangular Model (TM), which was initially introduced by Kulpa (1997). In the TM, a time interval *I*, which starts at I^- and ends at I^+, is represented by an intersection point *P* of two corresponding lines L_1 and L_2, as is shown in Figure 3.1. Thus, given a specific time interval, any sub interval inside it can be represented by a corresponding point in the TM. In addition to discrete time intervals, the TM was extended to the CTM by Qiang et al. (2014) to represent continuous temporal data. Thus, the CTM has the ability to represent the attribute values during all time intervals. The attribute value during a time interval can be calculated using an algebra operator (such as mean, maximum and minimum) based on the attributes at the finest sampled timestamps within this time interval, and the attribute values at the timestamps between any two neighbouring sampled timestamps can be derived using interpolation. Through colour-coding, the continuous field of the CTM can be displayed as an image, in which each colour corresponds to the attribute value of a specific time interval.

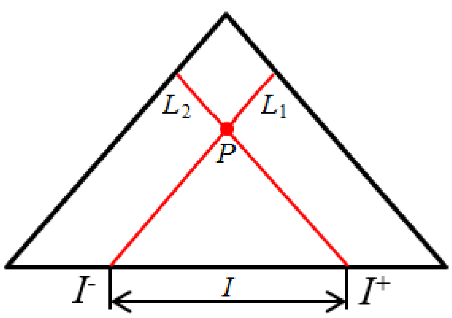


Figure 3.1. An illustration of the TM.

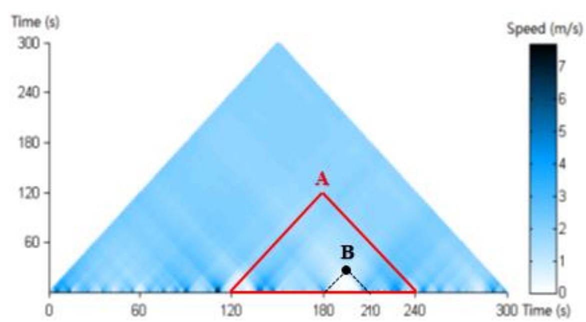


Figure 3.2. An illustration of the CTM.

Figure 3.2 shows an example of the CTM, which represents the values of the mean speed of a football player during the first five minutes. Obviously, the values of the mean speed during any time interval within [2, 4] minutes can be found in triangle *A*. For example, the value of point *B* corresponds to the mean speed of this player during the time interval [3, 3.5] minutes.

3.3 Introduction of the sequences

Numerous sequences can be generated in the CTM. However, not all of them are interesting. We construct four types of sequences that might have specific meanings. In the following, the basic concepts and the four types of sequences are introduced.

3.3.1 Basic concepts

3.3.1.1 Tevel

The basic element in the CTM is the **t**emporal **ev**olution **el**ement, or in short, tevel, which is named according to the term ‘stevel’ proposed by Van de Weghe et al. (2014). A tevel is a time interval represented as I_i^t, where *i* is the beginning of the time interval and *t* is the temporal scale. For example, the tevel I_2^4, which is denoted by the red point in Figure 3.3, means that the tevel begins at timestamp 2 and has a temporal scale of 4. Thus, the tevel equals to the time interval [2, 6]. If each tevel at the finest temporal scale is assigned a value, then the values for the tevels at coarser temporal scales can be calculated. The value of a tevel is represented as $V(I_i^t)$, where *i* and *t* have the same meanings as those in I_i^t.

3.3.1.2 Sequence

A sequence consists of a series of tevels. Assume there are *n* known timestamps (i.e., 0, 1, 2, …, *n*-1) at the finest temporal scale. All tevel values can form a maximum set, which can be represented as $S = \{V(I_i^t) |\ (0 \leq i \leq n-1) \wedge (0 \leq t \leq n-i)\}$. Any subset of S is considered as a sequence. Take Figure 3.4 for instance, the sequence S_1 consisting of eight tevels is represented as $S_1 = \{V(I_i^0) | i \in \mathbb{N} \wedge (0 \leq i \leq 7)\}$. Assume the values of the eight tevels are 3, 6, 1, 4, 6, 9, 7 and 5, respectively, then sequence S_1 can be visualised as a curve plotted in Figure 3.5.

3.3.2 The four types of sequences

3.3.2.1 Scaling-at sequence

The tevels involved in a sequence of this type have different beginnings of time intervals but the same temporal scale. Hence, a sequence of this type can be represented as

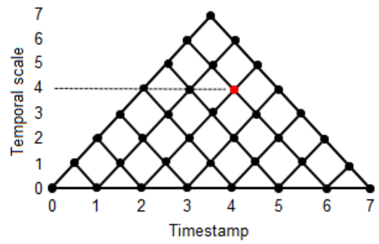


Figure 3.3. An illustration of tevel I_2^4.

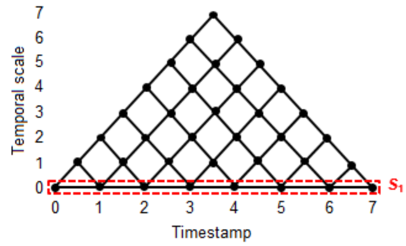


Figure 3.4. An illustration of sequence S_1.

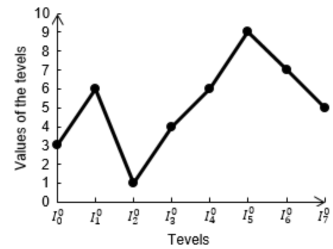


Figure 3.5. Visualisation of sequence S_1.

$S = \{V(I_i^k) | 0 \leq i \leq n-1\}$, where $n-1$ means the largest timestamp, k is a constant denoting the temporal scale and $0 \leq k \leq n-1$. Two such types of sequences S_1 and S_2 are illustrated in Figure 3.6(a). According to this type of sequences, the changes of motion attributes of moving objects across different temporal scales can be investigated.

3.3.2.2 Beginning-at sequence

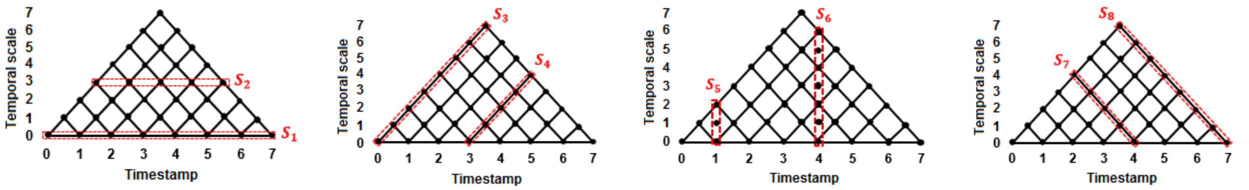
The tevels involved in a sequence of this type have the same beginning of time intervals but different temporal scales. Hence, a sequence of this type can be represented as $S = \{V(I_k^t) | \ 0 \leq t \leq n-1\}$, where $n-1$ means the largest timestamp, k is a constant denoting the beginning of time intervals and $0 \leq k \leq n-1$. Figure 3.6(b) gives an illustration of two such types of sequences S_3 and S_4.

3.3.2.3 Centring-at sequence

The tevels involved in a sequence of this type have the same centre of time intervals but different temporal scales. Hence, a sequence of this type can be represented as $S = \{V(I_{k-\frac{t}{2}}^t) | \ 0 \leq t \leq 2 * k\}$ when $0 \leq k \leq \frac{n-1}{2}$ and $S = \{V(I_{k-\frac{t}{2}}^t) | \ 0 \leq t \leq 2 * (n-1-k)\}$ when $\frac{n-1}{2} < k \leq n-1$, where $n-1$ denotes the largest timestamp, k is a constant denoting the centre of time intervals and $0 \leq k \leq n-1$. Figure 3.6(c) illustrates two such types of sequences S_5 and S_6.

3.3.2.4 Ending-at sequence

The tevels involved in a sequence of this type have the same end of time intervals but different temporal scales. Hence, a sequence of this type can be represented as $S = \{V(I_{k-t}^{t})|0 \leq t \leq k\}$, where $n-1$ denotes the largest timestamp, k is a constant denoting the end of time intervals and $0 \leq k \leq n-1$. Two such types of sequences S_7 and S_8 are shown in Figure 3.6(d).



(a) (b) (c) (d)

Figure 3.6. Illustrations of the four types of sequences: (a) scaling-at sequence ($S_1 = \{V(I_i^0)|i \in \mathbb{N} \wedge (0 \leq i \leq 7)\}$, $S_2 = \{V(I_i^3)|i \in \mathbb{N} \wedge (0 \leq i \leq 4)\}$), (b) beginning-at sequence ($S_3 = \{V(I_0^t)|t \in \mathbb{N} \wedge (0 \leq t \leq 7)\}, S_4 = \{V(I_3^t)|t \in \mathbb{N} \wedge (0 \leq t \leq 4)\}$), (c) centring-at sequence ($S_5 = \{V(I_{1-\frac{t}{2}}^t)|t \in \mathbb{N} \wedge (0 \leq t \leq 2)\}, S_5 = \{V(I_{4-\frac{t}{2}}^t)|t \in \mathbb{N} \wedge (0 \leq t \leq 6)\}$), and (d) ending-at sequence ($S_7 = \{V(I_{4-t}^t)|t \in \mathbb{N} \wedge (0 \leq t \leq 4)\}, S_8 = \{V(I_{7-t}^t)|t \in \mathbb{N} \wedge (0 \leq t \leq 7)\}$).

Among the four types of sequences, the latter three somewhat appear to be novel, since the tevels involved in the same sequence, on the one hand, are all at different temporal scales, and on the other hand, all have some property in common, e.g., sharing the same timestamp as the beginning/centre/end of the time intervals. Interesting information hidden in movement data might be revealed based on these four types of sequences.

3.4 Knowledge discovery based on sequence analysis

In this section, two research aims are derived based on the four types of sequences: (1) investigating the changes of motion attributes of moving objects across different temporal scales, and (2) detecting the time intervals during which active events might have occurred.

3.4.1 Investigating the changes of motion attributes across different temporal scales

The influences of temporal scales on moving objects can be explored by investigating the changes in motion attributes across different temporal scales. As this chapter mainly focuses on methodology, only the attribute of speed is taken as a representative, since speed is one of the attributes that has been frequently discussed in current research, such as in the work by Laube & Purves (2011), Chavoshi et al. (2014), Soleymani et al. (2014) and Zhang et al. (2016). Statistic measures can be employed to capture the distinctive features of the speed distribution. In this research, only the measure of mean, which denotes the average value in the speed distribution, is taken into consideration, as this can reveal interesting findings that have not been mentioned in previous research. For each moving object, the motion attributes with or without statistic measures can be visualised as a curve, according to which the changes across different temporal scales can be investigated.

In addition, we provide a simple yet useful method to give recommendations on the selection of optimal scales, at which specific demands might be met. The scale selection method stands on two measures, which can be used to measure the overall variations of a sequence: standard deviation and sinuosity (Dodge et al., 2012). Standard deviation reflects the changes of amplitude variation of a sequence, and sinuosity reflects the changes of frequency variation of a sequence. For each sequence, the corresponding values of standard deviation and sinuosity can be calculated. To compare the variations of different sequences at different temporal scales, the values of standard deviation and sinuosity have to be normalised. The normalisation method adopted is the Z-score method (Dodge et al., 2012). After normalisation, for each measure, three types of values exist: high values, neutral values and low values. Note that in theory neutral values mean that the values equal to zero, however, this cannot certainly appear in practice. Hence, a threshold ought to be set to distinguish neutral values. After analysing the values, we consider that 0.01 is an applicable threshold, which is the same as that used in (Dodge et al., 2012) as well. Thus, any value within [-0.01, 0.01] can be considered as a neutral value. The values larger than 0.01 are considered as high values, and those smaller than -0.01 as low values. Hence, based on the two measures and their different values, nine categories are derived and listed in Table 1. Figure 3.7 gives an illustration of the nine categories. According to Table 3.1, the variation of any sequence at any temporal scale belongs to a certain category from A to I. Hence, one can select the optimal scale(s) based on different demands on the extent of variations of a sequence in terms of this table.

Table 3.1. The classifications of the variations of a sequence.

Sinuosity	Standard deviation		
	High	Neutral	Low
High	A	B	C
Neutral	D	E	F
Low	G	H	I

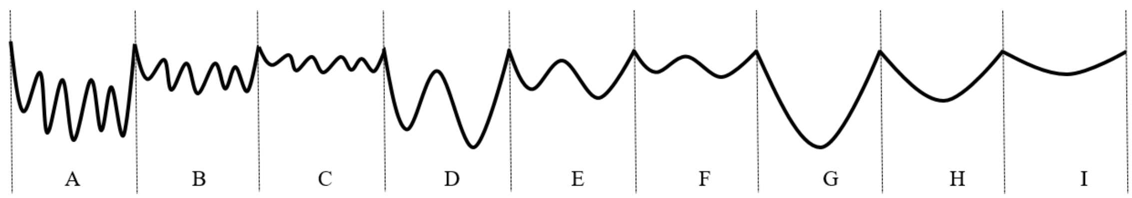


Figure 3.7. Illustration of the nine categories of the variations of a sequence.

3.4.2 Detecting the time intervals during which active events might have occurred

The detection of time intervals is of importance, since active or inactive events might occur during specific time intervals. Active events are considered as events happened during the time intervals which make the moving objects have large variations in geographical location. Scale is an important factor that should be taken into account when detecting such time intervals. However, traditionally, time intervals are detected based on a sequence at only one temporal scale. In this chapter, we propose an alternative approach to detect time intervals based on sequences at multiple temporal scales. Specifically, in this chapter, the time intervals are detected based on the three types of sequences: beginning-at sequence, centring-at sequence and ending-at sequence.

The approach consists of six steps. A simple dataset is used to illustrate how the approach works. Suppose this dataset corresponds to the speed of a moving object during the time interval [0, 7] minutes, and the selected temporal scales are 0 minute, 1 minute, 2 minutes, 3 minutes, 4 minutes, 5 minutes, 6 minutes and 7 minutes. Note that in order to make the units to be consistent, 1 second is replaced by 0 minute.

3.4.2.1 Selecting the type of sequence to be analysed

In this illustration, the beginning-at sequence is taken as an example.

3.4.2.2 Generating the sequences by determining all the corresponding tevels

The eight sequences are $\{S_{i+1} = \{V(I_i^t)|i \in \mathbb{N} \land (0 \leq t \leq 7)\}|t \in \mathbb{N} \land (0 \leq t \leq 7 - i)\}$ and are shown in Figure 3.8(a). Assume each tevel at the finest temporal scale is assigned a corresponding speed value (e.g., 1, 3, 7, 2, 4, 5, 3 and 2), the speed values of the other tevels can be calculated based on the CTM. The eight sequences thus can be visualised as curves that are presented in Figure 3.8(b). The following calculations are all based on the eight sequences and the corresponding speed values.

3.4.2.3 Calculating the standard deviation and sinuosity of each sequence

We propose to detect time intervals based on two important measures: standard deviation and sinuosity. They both can be used to measure the variations of a sequence. A small value indicates a small variation while a large value means a large variation. After calculating, the values of both measures are then normalised using the Z-Score method. The normalised values of the two measures are plotted as curves that are shown in Figure 3.8(c).

3.4.2.4 Extracting the sequences with relatively large variations

If a sequence has a relatively large variation, there are more possibilities for active events to be occurred during the time intervals related to this sequence. As a large standard deviation and a large sinuosity indicate a large variation, the extraction of sequences with a large standard deviation and a large sinuosity is the key in this step. The method to extract the sequences with relatively large variations is demonstrated as follows.

(1) For the curve of standard deviation (or sinuosity), calculate the absolute differences of the values between any two neighbouring points (assume the calculated results are stored in R_1);

(2) Given a parameter p_1 ($0 \leq p_1 \leq 1$), the value at the $(100^* p_1)^{\text{th}}$ percentile in R_1 is

considered as a threshold to determine whether a sequence has a large variation. If a value is larger than the threshold, the corresponding sequence is regarded as having a large variation, otherwise, not;
(3) For a point in the curve of standard deviation (or sinuosity), if the differences of its value and the values of its neighbouring points are both larger than the threshold, the sequence corresponding to this point is considered to be with a large variation.

According to this method, the corresponding sequences with large variations can be extracted based on either standard deviation or sinuosity. Note that only the sequences that are extracted according to both measures are considered as with large variations. For example, in the illustrated dataset, only the third sequence (i.e., S_3) is extracted and it is plotted in Figure 3.8(d).

3.4.2.5 Getting the time intervals based on the extracted sequences

If the speed increases more than a specific threshold during a time interval than that during its last time interval, it indicates that active events might have occurred during this time interval, thus resulting in the increase of speed. According to this, the time intervals during which active events might have occurred can be detected based on the method below.

(1) For a sequence, calculate the differences of the speed values of any two neighbouring points (assume the calculated results are stored in R_2);
(2) Given a parameter p_2 ($0 \le p_2 \le 1$), the value at the $(100^* p_2)^{th}$ percentile in R_2 is considered as a threshold to determine the potential time intervals;
(3) For each point in the sequence, if the differences of its speed value and that of its last point is larger than the threshold, the corresponding time interval between these two points are considered as the detected time interval.

For example, according to this method, the detected time intervals based on the sequence displayed in Figure 3.8(d) is [4, 5].

3.4.2.6 Validating the time intervals based on the corresponding dataset

Similarly, based on the centring-at sequence and the ending-at sequence, corresponding

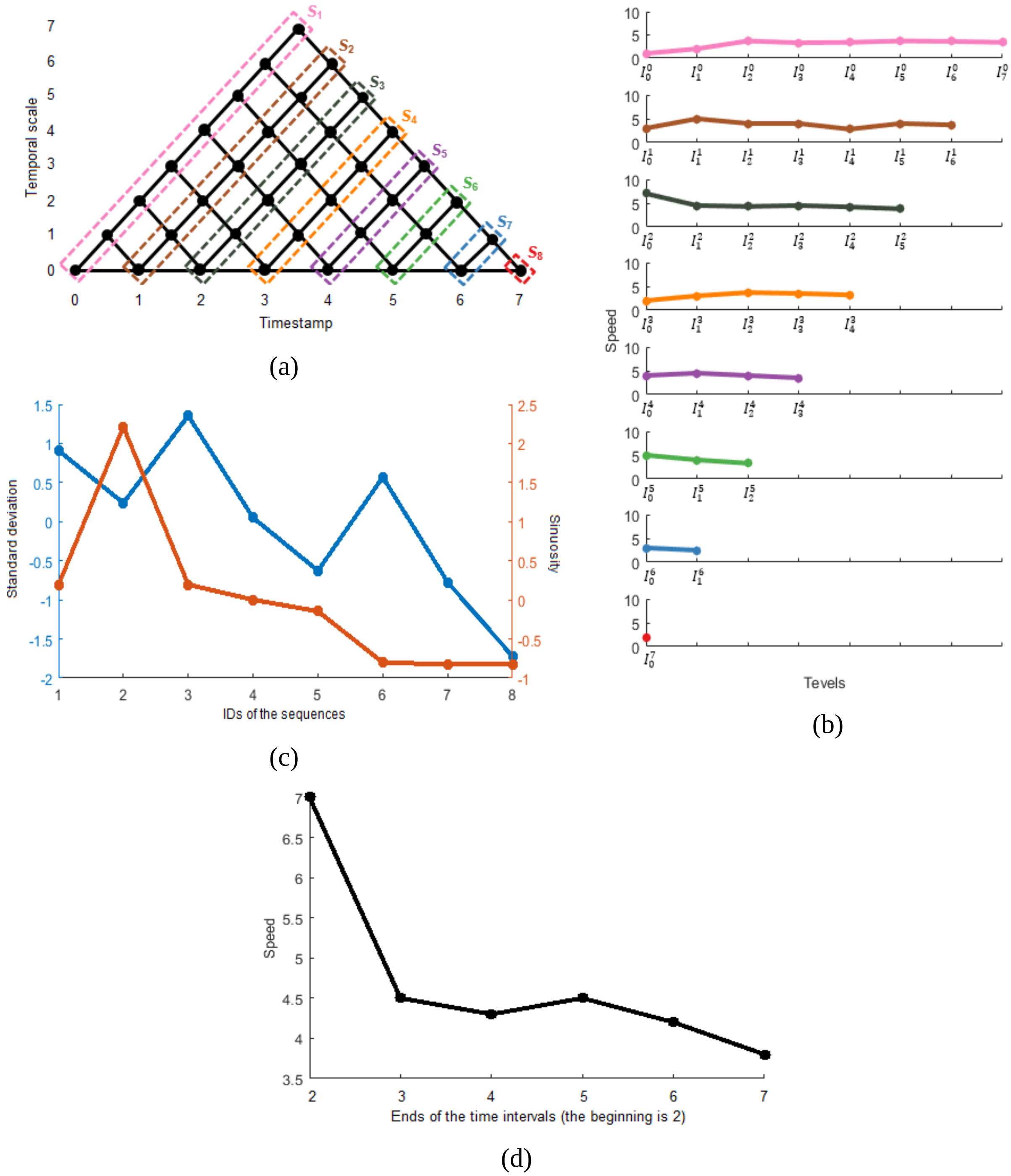


Figure 3.8. Illustrations of the proposed approach: (a) the eight sequences generated based on the CTM, (b) the visualisation of the eight sequences, (c) the plot of standard deviation and sinuosity (after normalisation), and (d) the plot of the sequence with relatively large variation.

time intervals can be detected. The final detected time intervals are considered as the union of all the time intervals detected respectively based on the three types of sequences. Finally, the time intervals have to be validated based on the corresponding dataset so that the accuracy can be evaluated.

3.5 Case study

3.5.1 Dataset

The dataset adopted in this case study is the movement data of football players, which were tracked during a real football match lasting about 95.5 minutes between ‘Club Brugge KV’ and ‘Standard de Liège’ in Belgium on the second of March 2014. For simplicity, we call the teams ‘Club Brugge’ and ‘Standard Liège’ respectively in the remainder of the text. The dataset includes both spatio-temporal information and semantic information. The spatio-temporal information mainly denotes the discrete points with a format of (*id*, *x*, *y*, *t*), where *id* identifies a specific player, *x* and *y* denote the *x* and *y* coordinates of the player’s position, and *t* corresponds to the timestamp. The semantic information used in this chapter is mainly the information of both teams (especially the basic information of the players, such as the names, the *id* numbers, and the positions) and the events that happened during the match (including the event name, the time of occurrence and the *id* of the actors). Given the vast volume of the dataset and according to the needs in the chapter, the temporal scale employed in this dataset is set to 1 second. Due to the limit of the chapter, in this case study, only the players of Club Brugge are taken into account. For the scaling-at sequence, the speed of the eight players who played for the whole match are used, since our aim is to investigate the motion attributes of individual players across different temporal scales. For the remaining three types of sequences, the speed of all the players (except the goalkeeper) are adopted to detect the time intervals during which active events might have occurred, as an event usually involves multiple players. With respect to the temporal scales, only the integral minutes are selected, namely the temporal scales are 0 minute (i.e., 1 second), 1 minute, 2 minutes, .., and 95 minutes, although any temporal scale within 1 second and 95.5 minutes is applicable in theory. Therefore, there are 96 sequences for each type. In detail, the 96 sequences for the scaling-at sequence are $\{S_{t+1} = \{V(I_i^t) | t \in \mathbb{N} \wedge (0 \leq t \leq$

$95)\}|i \in \mathbb{N} \wedge (0 \leq i \leq 95 - t)\}$, the 96 sequences for the beginning-at sequence are $\{S_{i+1} = \{V(I_i^t)|i \in \mathbb{N} \wedge (0 \leq t \leq 95)\}|t \in \mathbb{N} \wedge (0 \leq t \leq 95 - i)\}$, the 96 sequences for the centring-at sequence are $\{S_{i+1} = \{V(I_{i-\frac{t}{2}}^t)|i \in \mathbb{N} \wedge (0 \leq i \leq 47)\}|t \in \mathbb{N} \wedge (0 \leq t \leq 2 * i)\}$ and $\{S_{i+1} = \{V(I_{i-\frac{t}{2}}^t)|i \in \mathbb{N} \wedge (48 \leq i \leq 95)\}|t \in \mathbb{N} \wedge (0 \leq t \leq 2 * (n - 1 - i))\}$, and the 96 sequences for the ending-at sequences are $\{S_{i+1} = \{V(I_{i-t}^t)|i \in \mathbb{N} \wedge (0 \leq i \leq 95)\}|t \in \mathbb{N} \wedge (0 \leq t \leq i)\}$.

3.5.2 The changes in motion attributes across different temporal scales

The changes of speed of the eight players across the 96 temporal scales are visualised using the scatter plot, which can be seen in Figure 3.9. Note that 'temporal scale(s)' is replaced by 'scale(s)' for simplicity in the remainder of the text. In Figure 3.9, the speed values are all normalised to be within [0, 1] so that they can be qualitatively compared. In the scatter plot the horizontal axis denotes the scales and the vertical axis means the speed values, each of which is represented by a point.

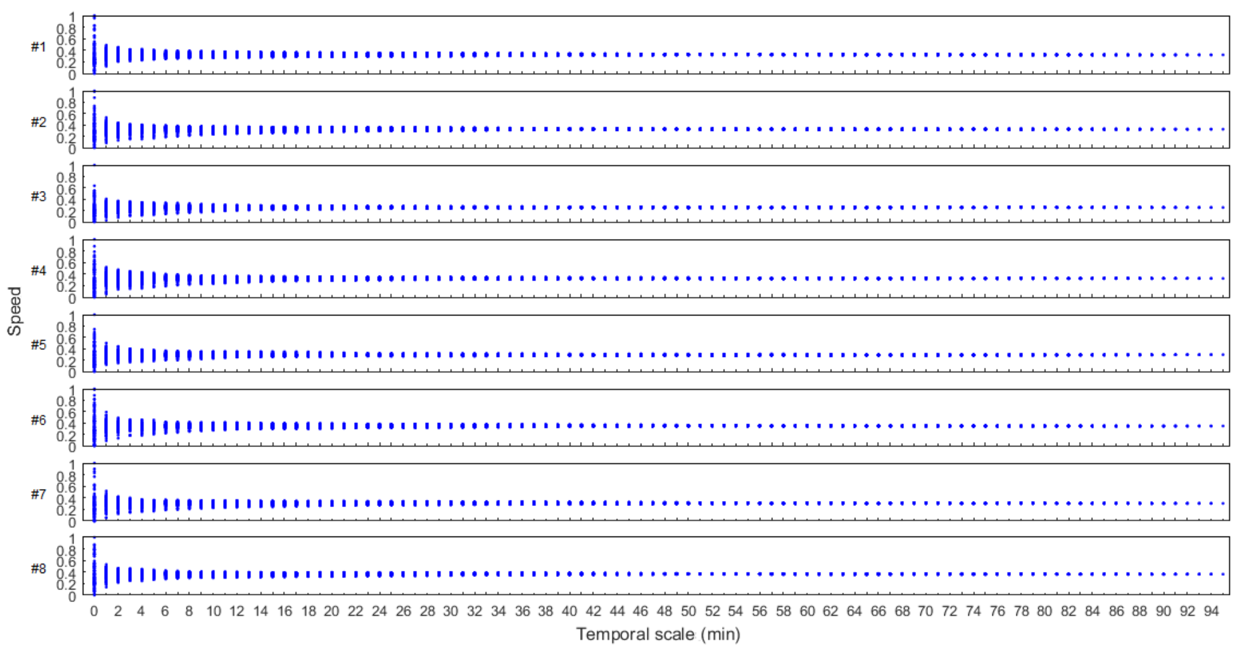


Figure 3.9. The changes of speed of the eight players across the 96 scales.

According to Figure 3.9, we can notice that the speed of the eight players shows a number of general features, which can be summarised as follows:

(1) For each player, the overall variance in speed decreases as the scale increases;

(2) For each player, the minimum value increases and the maximum value decreases as the scale increases;

(3) The eight players exhibit very similar overall patterns, and the patterns are unaffected to the roles of the individual players (note that players #1 ~ #4 are defenders, player #5 is a goalkeeper, and players #6 ~ #8 are midfielders).

The changes of the mean speed of the eight players across all scales are shown in Figure 3.10. Note that the variations of the speed values are all normalised to be within a range of 1 so that they can be qualitatively compared. As shown in Figure 3.10, on a whole, three different patterns can be found: pattern 1 (players #1 and #8), pattern 2 (players #2, #3, #4, #6 and #7) and pattern 3 (player #5), which are generalised and shown in Figure 3.11. According to Figure 3.11, we can notice that on the one hand, generally the changes of the mean speed for pattern 1 and pattern 3 are very obvious as the scale increases, but is relatively stable for pattern 2, and on the other hand, pattern 1 and pattern 3 look almost opposite. This demonstrates that different players may exhibit different patterns among all the scales, and the mean speed values tend to be relatively large (or small) around the median scale when patterns 1 (or 3) occurs. One common characteristic for the three patterns is that the changes (either increase or decrease) from the first scale to the second scale are dramatically large. This indicates that the mean speed might be sensitive to particularly small scales, e.g., scales less than 1 minute.

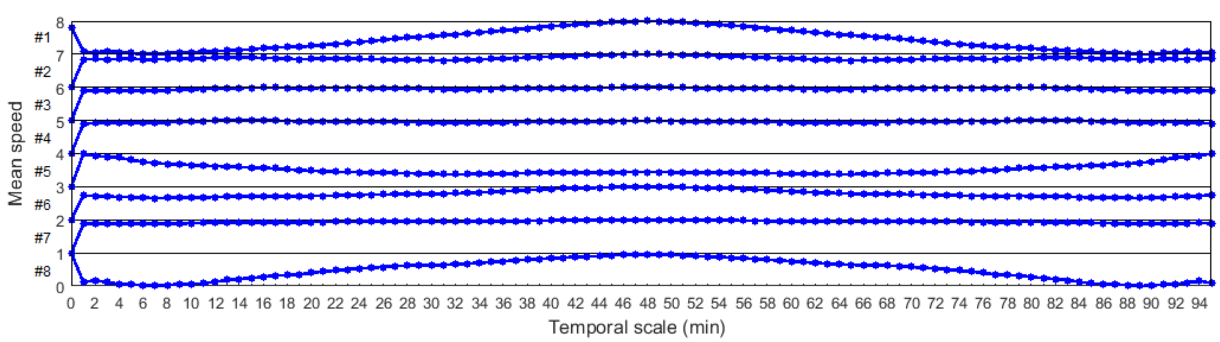
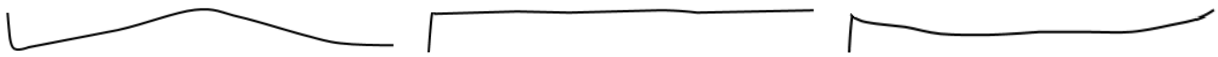


Figure 3.10. The changes of the mean speed of the eight players across the 96 scales.



(a) (b) (c)

Figure 3.11. The three generalised patterns: (a) pattern 1, (b) pattern2, and (c) pattern 3.

The findings based on the approach coincide well with those presented in (Laube & Purves, 2011) on a whole, although different types and numbers of moving objects are investigated. In addition, we have found three interesting patterns by taking the statistics measure of mean into consideration, which has not been conducted in former research. This demonstrates that the approach proposed in this chapter is valid and can be applied to analyse movement data from a cross-scale perspective.

According to the scale selection method, we draw a figure that demonstrates the applicable scales corresponding to the specific categories of sequence variations. As shown in Figure 3.12, only five categories listed in Table 3.1 appear among all the scales: categories A, B, C, F and I. Specifically, for players #1, #6 and #7, categories A, B, C, F and I appear; for players #3, #4 and #8, only categories A, C, F and I appear, and for players #2 and #5, only categories A, C and I appear. This demonstrates that, on the one hand, for each player, not all mentioned categories (listed in Table 3.1) can certainly appear among all the scales in practice, and on the other hand, different players usually exhibit different categories among all the scales, but some specific categories tend to appear certainly, such as categories A, C and I. Besides, by analysing categories A, C and I, we can find that they tend to appear at relatively fine-scale, intermediate-scale and coarse-scale, respectively. According to this figure, one can find out the optimal scales at which specific needs can be met. For example, if one is interested in the scales at which a sequence exhibits low standard deviation and high sinuosity (i.e., category C) for player #3, then any scale between 6 minutes and 22 minutes appears to be applicable.

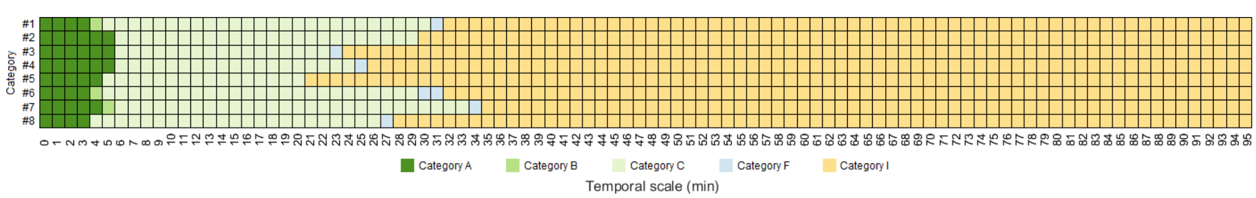


Figure 3.12. Recommendations of the optimal scales for the eight players at the 96 scales when analysing the mean speed.

3.5.3 The detection of time intervals during which active events might have occurred

The dataset used in this section is the speed of all the players (except the goalkeeper) of Club Brugge during the whole match, since an active event usually involves multiple players in a match, not just one specific player. In the adopted dataset, we consider the four events, i.e., 'shot not on target', 'shot on target', 'goal' and 'foul direct free-kick', as active events, and use them to validate the detected time intervals. This is because that among the events in the original dataset, these four have the most possibilities to result in rapid/obvious/instantaneous changes of speed. The basic duration of the time intervals to be detected is one minute.

Apparently, the final results depend on the parameter combinations adopted. Hence, numerous different parameter combinations were tested to compare the results. The results under nine typical parameter combinations are intuitively shown in Figure 3.13. Note that in Figure 3.13, the detected time intervals are marked either in blue or in black. Blue indicates that at least one active event happened during the corresponding time intervals, and black indicates that no event happened during the corresponding time intervals. Besides, the actual time intervals during which shot events (either 'shot not on target', 'shot on target' or 'goal') and 'foul direct free-kick' events happened are marked respectively in red and in green. According to Figure 3.13, the corresponding quantitative comparisons of the results are shown in Table 3.2. In Table 3.2, *N* denotes the number of target time intervals that are derived from the original dataset, *ND* denotes the number of time intervals that are detected under each parameter combination and *NCD* means the number of correct time intervals that are detected. Based on *N*, *ND* and *NCD*, two new indices, namely *PD* and *PCDD*, are derived in order to explore the accuracies of the results detected by the approach. *PD* and *PCDD* are respectively calculated as follows: $PD = ND/N$ and $PCDD = NCD/ND$. The changes of *PD* and *PCDD* with parameters p_1 and p_2 are plotted based on Table 3.2 and shown respectively in figures (a) and (b) in Figure 3.14. From Figure 3.14, we can notice that when one parameter is fixed, the values of *PD* decrease as the other parameter increases, but the values of *PCDD* increase as the other parameter increases. In our case, we propose that a good parameter combination should be one that can make *PD* and *PCDD* to be as large as possible, while the variations between *PD* and *PCDD* should be as small as possible. Therefore, the values of the mean and the standard deviation of *PD* and *PCDD* under each parameter combination are calculated, and they are shown in Table 3.3. Besides, the values of the difference between the mean and the standard deviation, which can be denoted as the potential minimum accuracy, are calculated and listed in Table 3.3 as well. According to Table 3.3, $p_1 = 0.75$ and $p_2 = 0.50$ appears to be a good parameter combination.

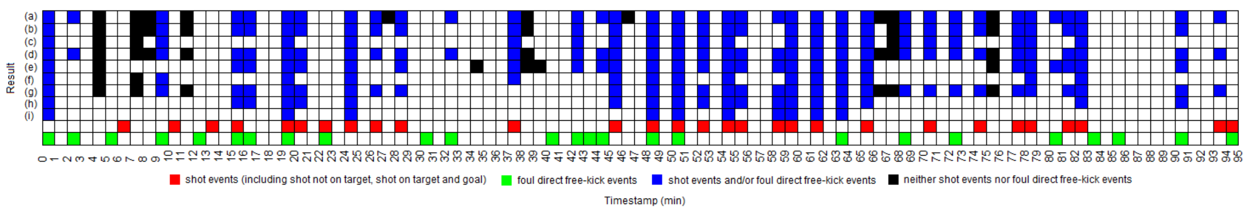


Figure 3.13. The results under nine different parameter combinations: (a) $p_1 = 0.50$, $p_2 = 0.50$, (b) $p_1 = 0.50$, $p_2 = 0.75$, (c) $p_1 = 0.50$, $p_2 = 0.90$, (d) $p_1 = 0.75$, $p_2 = 0.50$, (e) $p_1 = 0.75$, $p_2 = 0.75$, (f) $p_1 = 0.75$, $p_2 = 0.90$, (g) $p_1 = 0.90$, $p_2 = 0.50$, (h) $p_1 = 0.90$, $p_2 = 0.75$, and (i) $p_1 = 0.90$, $p_2 = 0.90$.

Table 3.2. The quantitative comparisons of the results under different parameter combinations.

Parameter combination	***N***	***ND***	***NCD***	***PD***	***PCDD***
$p_1 = 0.50$, $p_2 = 0.50$	48	46	36	0.958	0.783
$p_1 = 0.50$, $p_2 = 0.75$	48	37	30	0.771	0.811
$p_1 = 0.50$, $p_2 = 0.90$	48	24	21	0.500	0.875
$p_1 = 0.75$, $p_2 = 0.50$	48	43	35	0.896	0.814
$p_1 = 0.75$, $p_2 = 0.75$	48	32	27	0.667	0.844
$p_1 = 0.75$, $p_2 = 0.90$	48	17	15	0.354	0.882
$p_1 = 0.90$, $p_2 = 0.50$	48	36	30	0.750	0.833
$p_1 = 0.90$, $p_2 = 0.75$	48	19	19	0.396	1.000
$p_1 = 0.90$, $p_2 = 0.90$	48	10	10	0.208	1.000

Note: '*N*', '*ND*', '*NCD*', '*PD*' and '*PCDD*' respectively denote 'number of target time intervals', 'number of the detected time intervals', 'number of the correctly detected time intervals', 'percentage of the detected time intervals' and 'percentage of the correctly detected time intervals among all the detected time intervals'.

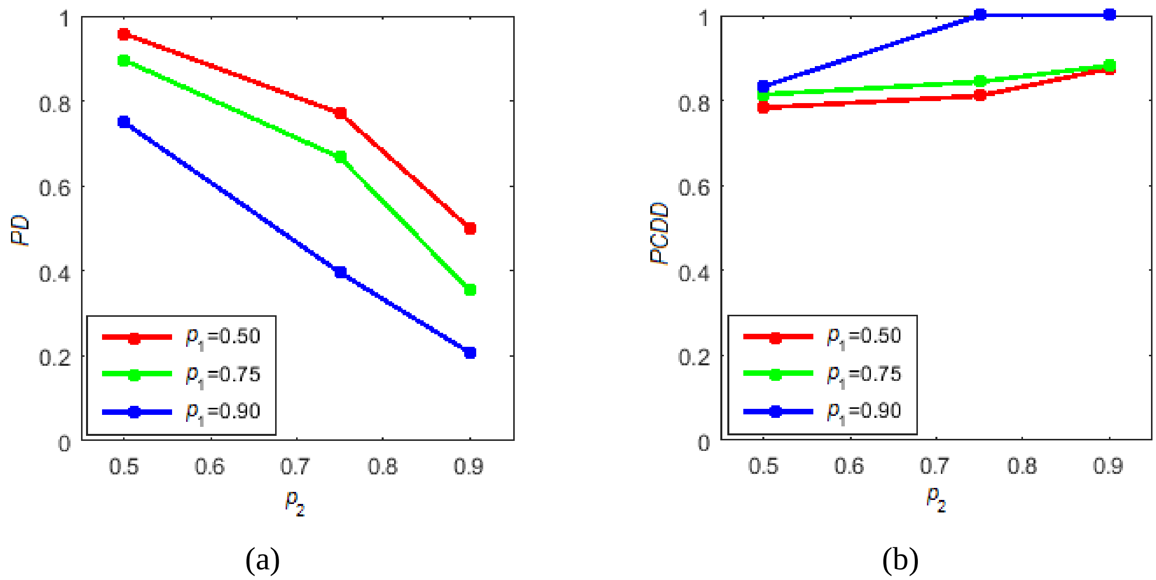


Figure 3.14. The changes of *PD* and *PCDD* with the parameters p_1 and p_2: (a) *PD*, and (b) *PCDD*.

Table 3.3. The values of the mean, the standard deviation and the difference between the mean and the standard deviation of *PD* and *PCDD* under different parameter combinations.

Parameter combination	Mean	Standard deviation	Mean-Standard deviation
$p_1 = 0.50$, $p_2 = 0.50$	0.871	0.124	0.747
$p_1 = 0.50$, $p_2 = 0.75$	0.791	0.028	0.763
$p_1 = 0.50$, $p_2 = 0.90$	0.688	0.265	0.423
$p_1 = 0.75$, $p_2 = 0.50$	0.855	0.058	0.797
$p_1 = 0.75$, $p_2 = 0.75$	0.756	0.125	0.631
$p_1 = 0.75$, $p_2 = 0.90$	0.618	0.373	0.245
$p_1 = 0.90$, $p_2 = 0.50$	0.792	0.059	0.733
$p_1 = 0.90$, $p_2 = 0.75$	0.698	0.427	0.271
$p_1 = 0.90$, $p_2 = 0.90$	0.604	0.560	0.044

In order to compare the pros and cons of the proposed approach, we also show the time intervals that are detected with traditional approaches which do not take cross-scale into account, namely the time intervals detected based on sequences of a single scale. In this case, we take two single scales, i.e., 1 second and 1 minute, as an example. The results are presented in Figure 3.15 (scale: 1 second) and Figure 3.16 (scale: 1 minute), respectively. The corresponding quantitative comparisons are listed in Table 3.4. Based on Table 3.4, we can notice that: (1) the values of *ND* at the scale of 1 second are much larger than *N*, which makes the results to be meaningless to some extent; (2) although the values of *ND* at the scale of 1 minute appear reasonable, the values of *PCDD* are relatively small, which makes the accuracy of the results not that high, and (3) according to Table 3.4, we can infer that the parameter of p = 0.50 at the scale of 1 minute is the optimal parameter, however, the overall accuracy is still lower than that based on the proposed approach. Therefore, the proposed approach which takes cross-scale into account has advantages and is effective in detecting the time intervals during which active events might have occurred with relatively high accuracy.

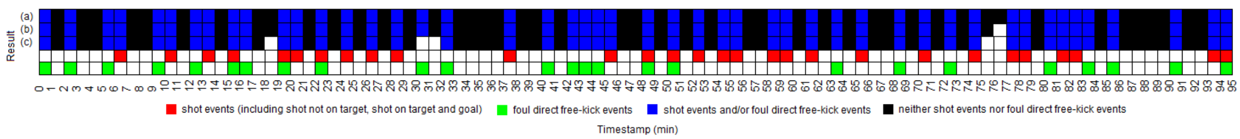


Figure 3.15. The results under different parameter combinations based on the sequences with a scale of 1 second: (a) p = 0.50, (b) p = 0.75, and (c) p = 0.90.

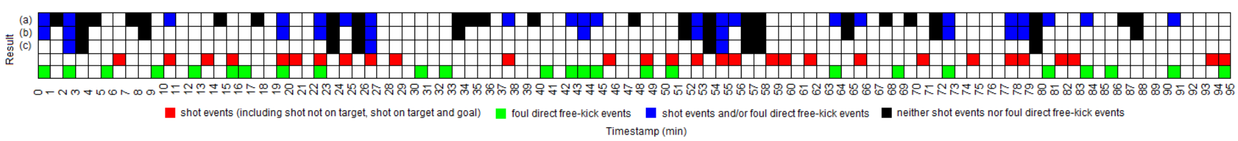


Figure 3.16. The results under different parameter combinations based on the sequences with a scale of 1 minute: (a) p = 0.50, (b) p = 0.75, and (c) p = 0.90.

Table 3.4. The quantitative comparisons of the results under different parameters based on the two single scales.

Parameter combination	*N*	*ND*	*NCD*	*PD*	*PCDD*
p = 0.50 (scale: 1 second)	48	95	48	1.979	0.505
p = 0.75 (scale: 1 second)	48	94	48	1.958	0.511
p = 0.90 (scale: 1 second)	48	90	47	1.875	0.522
p = 0.50 (scale: 1 minute)	48	47	22	0.979	0.468
p = 0.75 (scale: 1 minute)	48	24	11	0.500	0.458
p = 0.90 (scale: 1 minute)	48	10	3	0.208	0.300

Note: '*N*', '*ND*', '*NCD*', '*PD*' and '*PCDD*' respectively denote 'number of target time intervals', 'number of the detected time intervals', 'number of the correctly detected time intervals', 'percentage of the detected time intervals' and 'percentage of the correctly detected time intervals among all the detected time intervals'.

3.6 Conclusions and future work

Knowledge discovery in movement data is an important issue in many research domains.

Current research mainly focuses on one single scale. In this chapter, we propose a cross-scale oriented sequence analysis approach to discover knowledge in movement data. Key to this approach are the four different types of sequences that are constructed on the basis of the Continuous Triangular Model (CTM), a useful tool for representing data at multiple temporal scales. The approach mainly serves two specific aims: investigating the changes in motion attributes across different temporal scales and detecting the time intervals during which active events might have occurred. The changes of motion attributes of moving objects across different temporal scales are investigated based on the first type of sequences. Motion attributes include speed, distance, motion azimuth, and so on. As our aim is primarily methodological, only speed is investigated. The speed is investigated based on the movement data obtained from an entire football match. When investigating the changes of speed, the eight players of Club Brugge who played for the whole match are taken into account. Besides, the changes of speed involving a commonly used statistics measure are also investigated, according to which one is able to gain more abundant information. The time intervals during which active events might have occurred are detected based on the remaining three types of sequences. The data used for detecting the time intervals are the mean speed of all the players (except the goalkeeper) of Club Brugge during the whole match. By combining the corresponding time intervals detected based on each of the three types of sequences, the final detected time intervals can be acquired. The results show that the time intervals can be detected with high accuracies under suitable parameter combinations. The results acquired using the proposed approach are also compared with those that do not take cross-scale into account. The comparison demonstrates that the results acquired using the proposed approach have relatively high accuracies. Hence, the proposed approach is efficient in discovering useful knowledge in movement data.

In this chapter, only the attribute of speed is investigated. As part of future work, other attributes, either meaningful for all kinds of moving objects, such as distance and motion azimuth, or only meaningful for specific kinds of moving objects, e.g., ball possession of sports-related moving objects, can also be explored using this approach. Although for better comparison purposes, only the eight players who played for the whole match are taken into consideration, the players who played for only part of the match can also be taken for analysis when necessary. In addition, the proposed approach can be further extended to other application domains. For example, in the transportation domain, the time intervals during which the cars have a low speed can be detected using the proposed approach. This is of significant importance in practice since these time intervals are the potential periods during which traffic congestion has happened. Another interesting application domain, for example, might be animal movement studies.

References

Abbott, A. (1990). A primer on sequence methods. *Organization Science*, *1*(4), 375-392.

Andersson, M., Gudmundsson, J., Laube, P., & Wolle, T. (2008). Reporting leaders and followers among trajectories of moving point objects. *GeoInformatica*, *12*(4), 497-528.

Andrienko, G., Andrienko, N., Burch, M., & Weiskopf, D. (2012). Visual analytics methodology for eye movement studies. *IEEE Transactions on Visualization and Computer Graphics*, *18*(12), 2889-2898.

Buchin, M., Dodge, S., & Speckmann, B. (2012). Context-aware similarity of trajectories. *Lecture Notes in Computer Science*, *7478*, 43-56.

Buchin, M., Dodge, S., & Speckmann, B. (2014). Similarity of trajectories taking into account geographic context. *Journal of Spatial Information Science*, *9*, 101-124.

Chavoshi, S. H., De Baets, B., Neutens, T., De Tré, G., & Van de Weghe, N. (2015). Exploring dace movement data using sequence alignment methods. *PloS one*, *10*(7), e0132452.

Civilis, A., Jensen, C. S., & Pakalnis, S. (2005). Techniques for efficient road-network-based tracking of moving objects. *IEEE Transactions on Knowledge and Data Engineering*, *17*(5), 698-712.

Çöltekin, A., Fabrikant, S. I., & Lacayo, M. (2010). Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings.

International Journal of Geographical Information Science, *24*(10), 1559-1575.

Delafontaine, M., Versichele, M., Neutens, T., & Van de Weghe, N. (2012). Analysing spatiotemporal sequences in Bluetooth tracking data. *Applied Geography*, *34*, 659-668.

Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Van de Weghe, N., ... & Weibel, R. (2015). Analysis and visualisation of movement: an interdisciplinary review. *Movement ecology*, *3*(1), 5.

Dodge, S, Laube, P., & Weibel, R. (2012). Movement similarity assessment using symbolic representation of trajectories. *International Journal of Geographical Information Science*, *26*(9), 1563-1588.

Elragal, A., & EL-Gendy, N. (2013). Trajectory data mining: integrating semantics. *Journal of Enterprise Information Management*, *26*(5), 516-535.

Elsner, J. B., & Kara, A. B. (1999). *Hurricanes of the North Atlantic: climate and society*. Oxford University Press.

Gomez, G., López, P. H., Link, D., & Eskofier, B. (2014). Tracking of ball and players in beach volleyball videos. *PloS one*, *9*(11), e111730.

Goodchild, M. F. (2011). Scale in GIS: an overview. *Geomorphology*, *130*(1), 5-9.

Gudmundsson, J., van Kreveld, M., & Speckmann, B. (2007). Efficient detection of patterns in 2D trajectories of moving points. *Geoinformatica*, *11*(2), 195-215.

Gudmundsson, J., & Wolle, T. (2014). Football analysis using spatio-temporal tools. *Computers, Environment and Urban Systems*, *47*, 16-27.

Han, J., & Kamber, M. (2006). *Data mining: concepts and techniques*. Morgan kaufmann.

Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(1), 4-37.

Jeung, H., Yiu, M. L., Zhou, X., Jensen, C. S., & Shen, H. T. (2008). Discovery of convoys in trajectory databases. In *Proceedings of the VLDB Endowment* (pp. 1068-1080).

Kulpa, Z. (1997). Diagrammatic representation of interval space in proving theorems about interval relations. *Reliable Computing*, *3*(3), 209-217.

Laube, P., Imfeld, S., & Weibel, R. (2005). Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science*, *19*(6), 639-668.

Laube, P., & Purves, R. (2011). How fast is a cow? Cross-scale analysis of movement data. *Transactions in GIS, 15*(3), 401-418.

Lee, J. G., Han, J., & Li, X. (2008). Trajectory outlier detection: a partition-and-detect framework. In *Proceedings of the International Conference on Data Mining* (pp. 140-149).

Lee, J. G., Han, J., & Whang, K. (2007). Trajectory clustering, a partition-and-group framework. In *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data* (pp. 593-604).

Liu, L., Qiao, S., Zhang, Y., & Hu, J. (2012). An efficient outlying trajectories mining approach based on relative distance. *International Journal of Geographical Information Science*, *26*(10), 1789-1810.

Postlethwaite, C. M., Brown, P., & Dennis, T. E. (2013). A new multi-scale measure for analyzing animal movement data. *Journal of Theoretical Biology*, *317*, 175-185.

Qiang, Y., Chavoshi, S. H., Logghe, S., De Maeyer, P., & Van de Weghe, N. (2014). Multi-scale analysis of linear data in a two-dimensional space. *Information Visualization*, *13*(3), 248-265.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, *26*(1), 43-49.

Shamoun-Baranes, J., van Loon, E. E., Purves, R. S., Speckmann, B., Weiskopf, D., & Camphuysen, C. J. (2012). Analysis and visualization of animal movement. *Biology Letters*, *8*(1), 6-9.

Soleymani, A., Cachat, J., Robinson, K., Dodge, S., Kalueff, A., & Weibel, R. (2014). Integrating cross-scale analysis in the spatial and temporal domains for classification of behavioral movement. *Journal of Spatial Information Science*, *8*(8), 1-25.

Tappert, C. C., Suen, C. Y., & Wakahara, T. (1990). The state of the art in online handwriting recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *12*(8), 787-808.

Teimouri, M., Indahl, U. G., & Tveite, H. (2016). A method to detect inactive periods in animal movement using density-based clustering. *Applied Geography*, *73*, 102-112.

Turdukulov, U., Calderon Romero, A. O., Huisman, O., & Retsios, V. (2014). Visual mining of moving flock patterns in large spatio-temporal data sets using a frequent pattern approach. *International Journal of Geographical Information Science*, *28*(10), 2013-2029.

van Oosterom, P., & Stoter, J. (2010). 5D data modelling: full integration of 2D/3D space, time and scale dimensions. *Lecture Notes in Computer Science*, *6292*, 310-324.

Van de Weghe, N., De Roo, B., Qiang, Y., Versichele, M., Neutens, T., & De Maeyer, P. (2014). The continuous spatio-temporal model (CSTM) as an exhaustive framework for multi-scale spatio-temporal analysis. *International Journal of Geographical Information Science*, *28*(5), 1047-1060.

Vogel, M., Hamon, R., Lozenguez, G., Merchez, L., Abry, P., Barnier, J., ... & Robardet, C. (2014). From bicycle sharing system movements to users: a typology of Vélo' cyclists in Lyon based on large-scale behavioural dataset. *Journal of Transport Geography*, *41*, 280-291.

Wachowicz, M., Ong, R., Renso, C., & Nanni, M. (2011). Finding moving flock patterns among pedestrians through collective coherence. *International Journal of Geographical Information Science*, *25*(11), 1849-1864.

Xu, Z., Sandrasegaran, K., Kong, X., Zhu, X., Zhao, J., Hu, B., & Lin, C. C. (2013). Pedestrian monitoring system using Wi-Fi technology and RSSI based localization. *International Journal of Wireless & Mobile Networks*, *5*(4), 17-34.

Yuan, Y., & Raubal, M. (2014). Measuring similarity of mobile phone user trajectories-a Spatio-temporal Edit Distance method. *International Journal of Geographical Information Science*, *28*(3), 496-520.

Zhang, P., Beernaerts, J., Zhang, L., & Van de Weghe, N. (2016). Visual exploration of match performance based on football movement data using the Continuous Triangular Model. *Applied Geography*, *76*, 1-13.

Zhang, P., Beernaerts, J., & Van de Weghe, N. (2018). A hybrid approach combining the

Multi-Temporal Scale Spatio-Temporal Network with the Continuous Triangular Model for exploring dynamic interactions in movement data: a case study of football. *ISPRS International Journal of Geo-Information*, *7*(1), 31.

Zhang, P., Deng, M., & Van de Weghe, N., 2014. Clustering spatio-temporal trajectories based on kernel density estimation. *Lecture Notes in Computer Science*, 8579, 298-311.

4

A Hybrid Approach for Exploring Dynamic Interactions in Movement Data

Modified from: Zhang, P., Beernaerts, J., & Van de Weghe, N. (2018). A hybrid approach combining the Multi-Temporal Scale Spatio-Temporal Network with the Continuous Triangular Model for exploring dynamic interactions in movement data: a case study of football. *ISPRS International Journal of Geo-Information*, *7*(1), 31.

Abstract: Benefiting from recent advantages in location-aware technologies, movement data are becoming ubiquitous. Hence, numerous research topics with respect to movement data have been undertaken. Yet, the research of dynamic interactions in movement data is still in its infancy. In this chapter, we propose a hybrid approach combining the Multi-Temporal Scale Spatio-Temporal Network (MTSSTN) and the Continuous Triangular Model (CTM) for exploring dynamic interactions in movement data. The approach mainly includes four steps: first, the Relative Trajectory Calculus (RTC) is used to derive three types of interaction patterns; second, for each interaction pattern, a corresponding MTSSTN is generated; third, for each MTSSTN, the interaction intensity measures and three centrality measures (i.e., degree, betweenness and closeness) are calculated; finally, the results are visualised at multiple temporal scales using the CTM and analysed based on the generated CTM diagrams. Based on the proposed approach, three distinctive aims can be achieved for each interaction pattern at multiple temporal scales: (1) exploring the

interaction intensities between any two individuals; (2) exploring the interaction intensities among multiple individuals, and (3) exploring the importance of each individual and identifying the most important individuals. The movement data obtained from a real football match are used as a case study to validate the effectiveness of the proposed approach. The results demonstrate that the proposed approach is useful in exploring dynamic interactions in football movement data and discovering insightful information.

4.1 Introduction

With the technical developments in location-aware technologies such as GPS (global positioning system), RFID (radio-frequency identification), WiFi, Bluetooth, and image recognition, the position changes of moving objects over time can be tracked more easily than ever before. This has caused a proliferation of rich and voluminous movement data. Specific types of movement data include transportation related movement data (Civilis et al., 2005; Delafontaine et al., 2012; Xu et al., 2013; Vogel et al., 2014; Zhang et al., 2017), animal movement data (Laube et al., 2005; Shamoun-Baranes et al., 2012; Demšar et al., 2015), eye movement data (Andrienko et al., 2012), sports movement data (Gudmundsson & Wolle, 2014; Gomez et al., 2014; Zhang et al., 2016), as well as natural phenomena movement data (Lee et al., 2007). Benefiting from the large amount of tracked movement data, the analysis of movement data has become a state-of-the-art research theme in the community of geographical information science (GIScience). Currently, numerous methods to analyse movement data, such as movement pattern mining (Laube et al., 2005; Meijles, et al., 2014; Wang et al., 2015), movement visualisation (Andrienko et al., 2007; Andrienko et al., 2010; Kveladze, et al., 2015) and movement modelling (Hornsby & Egenhofer, 2002; Ahearn et al., 2016; Wang et al., 2016), have been undertaken extensively. In addition, the study of interactions in movement data has become active recently, but is still in its infancy (Long & Nelson, 2013; Long, 2015). In this chapter, we mainly focus on exploring the interaction issues in movement data aiming to provide additional insights into this relatively new research topic.

Moving objects commonly move in geographical space, in which the geographical context (e.g., the environments where moving objects live) is considered one of the important components. Therefore, the interactions in movement data can be categorised as the interactions between geographical contexts, the interactions between moving objects and geographical contexts, and the interactions between moving objects themselves. In this chapter, we only consider the interactions between/among moving objects themselves. Interactions can be classified as static interactions or dynamic interactions (Doncaster, 1990). In spatio-temporal data (e.g., movement data), static interactions are purely described by spatial properties, without taking account of the possibility of temporal avoidance or attraction between individuals, while dynamic interactions are defined based on both spatial and temporal components (Miller, 2015). Hence, ‘dynamic interaction’ is sometimes synonymously termed as ‘spatio-temporal interaction’ (Long, 2015). We mainly focus on dynamic interactions in this chapter.

Generally, dynamic interactions can be defined as the way the movements of individuals are related or the inter-dependency in the movements of individuals. For example, attraction and avoidance are two typical kinds of dynamic interactions (Miller, 2015). Typical research on dynamic interactions in movement data within the domain of GIScience is listed as follows. Miller (2012) analysed the dynamic interactions between individuals based on the GPS data of animals using five different techniques, thereby comparing the results acquired by the different techniques. Later, a null model approach (Miller, 2015) was developed by the same author to compare six dynamic interaction metrics using data on five brown hyena dyads in Northern Botswana. The comparisons highlighted the need for further study of appropriate methods for measuring and interpreting dynamic interactions (Miller, 2015). Long & Nelson (2013) introduced a new method, Dynamic Interactions (termed DI), for measuring dynamic interactions between pairs of moving objects. Six simulated datasets and two applied examples (i.e., team sports and wildlife) were used to validate the DI method. The results showed that the DI method was able to be used to measure dynamic interactions in movement data. Long et al. (2014) executed an examination of eight currently available indices of dynamic interactions in wildlife telemetry studies and compared the effectiveness of the indices. Long (2015) examined the statistical properties of a suite of currently available methods in dynamic interactions. In this work, the ability of each method in characterising and capturing different patterns of dynamic interactions was examined in practice. Konzack et al. (2017) proposed a new approach to analyse interactions between two trajectories and developed a prototype visual analytics tool to evaluate the approach based on three datasets.

By summarising the aforementioned studies, we can find that the current research mainly focuses on either comparing or evaluating existing interaction methods based on various datasets, or developing new methods to measure dynamic interactions between a pair of trajectories. Besides, most of the current research on dynamic interactions has been done at a single temporal scale. Few work has focused on exploring the dynamic interactions among multiple moving objects and at multiple temporal scales. In addition, few has aimed at exploring the importance of each individual and identifying the individuals which play relatively important roles in maintaining specific types of interaction patterns. Hence, in order to achieve the above-mentioned explorations, we develop a hybrid approach combining the Multi-Temporal Scale Spatio-Temporal Network (MTSSTN) and the Continuous Triangular Model (CTM). Currently, networks are widely used to explore many systems. However, in the data-rich era, it appears necessary to deal with data that are temporally evolving. To this end, temporal networks are proposed and applied to various domains (Lee et al., 2012; He & Chen, 2015; Holme, 2015). More recently, spatio-temporal networks have been proposed to enhance the abilities of networks (Von Landesberger et al., 2016; Williams & Musolesi, 2016). In this chapter, we propose a novel spatio-temporal network (i.e., Multi-Temporal Scale Spatio-Temporal Network) to extend the analytics applicability of spatio-temporal networks at multiple temporal scales. We then integrate the MTSSTN with the CTM to compose our proposed approach. In this approach, first, the Relative Trajectory Calculus (RTC) (Van de Weghe, 2004) is employed to derive specific types of interaction patterns. Second, the MTSSTN is generated based on a specific interaction pattern. Third, for each MTSSTN, the interaction intensity measures and three frequently used centrality measures in network theory (i.e., degree, betweenness and closeness) (Newman, 2003; Jiang & Claramunt, 2004) are calculated. Finally, the results are visualised using a multi-temporal scale visualisation tool, the CTM (Qiang et al., 2014; Zhang et al., 2016), and the results are analysed based on the generated CTM diagrams. Based on the generated CTM diagrams, the following aims can be achieved for each interaction pattern at multiple temporal scales: (1) exploring the interaction intensities between any two individuals; (2) exploring the general interaction intensities among multiple individuals, and (3) exploring the importance of each individual and identifying the most important individuals by applying map algebra operations to multiple CTM diagrams.

The remainder of this chapter is organised as follows. Section 4.2 gives a brief introduction to network theory, RTC and CTM. In section 4.3, the methodology of the proposed approach is described in detail. In section 4.4, a case study of football is conducted using the proposed approach, and the results are analysed. Some advantages and disadvantages of the proposed approach are discussed in Section 4.5. Finally, in section 4.6, the conclusions and recommendations for future work are described.

4.2 Background knowledge

In this section, the basic information on network theory, the Relative Trajectory Calculus (RTC) and the Continuous Triangular Model (CTM) are introduced, so that one can be able to understand the proposed approach easily.

4.2.1 Network theory

A large number of systems, either natural or man-made, are structured in the form of networks. Therefore, network theory has been widely used to explore the complex systems existing in many domains. Typical examples include social networks, World Wide Web, transportation networks, academic cooperation networks, biological networks, and so forth (Newman, 2003; Barrat et al., 2004). Essentially, a network is a graph which can be represented as $G = (V,\ E)$, where $V = \{v_1, v_2, \dots, v_n\}$ and $E = \{v_i v_j\}(1 \le i \le n, 1 \le$

$j \leq n$) respectively denote the set of vertices and edges. A graph can be connected or disconnected, directed or undirected, and weighted or unweighted. A connected, undirected and unweighted graph with *n* vertices can be represented by an adjacency matrix $R(G)$ as denoted by equation (4.1):

$$R(G) = \left[r(v_i, v_j)\right]_{n\times n}, \text{where } r(v_i, v_j) = \begin{cases} 1, & \textit{if vertices } v_i \textit{ and } v_j \textit{ are connected} \\ 0, & \textit{if vertices } v_i \textit{ and } v_j \textit{ are disconnected} \end{cases} \tag{4.1}$$

Note that in this case, matrix $R(G)$ is symmetric. Besides, all diagonal elements of $R(G)$ are zero. Hence, in practice, we can just calculate either the top right or the bottom left part of $R(G)$ in order to reduce the complexity. The other part of $R(G)$ can be computed via a symmetric transformation.

Several measures have been proposed to characterise the topological structural properties of a network or to investigate the importance of vertices. Among these measures, centrality measures are widely used. Frequently used centrality measures are degree, betweenness and closeness. In a graph, the degree of a vertex corresponds to the number of vertices that are directly connected to this vertex. Formally, the degree of a given vertex v_i $(1 \leq i \leq n)$ is calculated as follows:

$$C_D(v_i) = \sum_{j=1}^{n} r(v_i, v_j) \tag{4.2}$$

where C_D represents the degree and *n* is the total number of vertices in the graph.

The betweenness of a vertex measures to what extent the vertex is located in between the paths that connect pairs of vertices. Formally, the betweenness of a given vertex v_i $(1 \leq i \leq n)$ is calculated as follows:

$$C_B(v_i) = \frac{\sum_{j=1}^{n-1} \sum_{k=j+1}^{n} p_{v_j v_i v_k}}{\sum_{j=1}^{n-1} \sum_{k=j+1}^{n} p_{v_j v_k}} \tag{4.3}$$

where C_B denotes the betweenness, $p_{v_j v_i v_k}$ is the number of shortest paths from vertex v_j to vertex v_k that pass through vertex v_i, $p_{v_j v_k}$ represents the number of shortest paths from vertex v_j to vertex v_k, and *n* is the total number of vertices in the graph.

The closeness of a vertex measures the closeness of a vertex to all other vertices in a graph. Formally, the closeness of a vertex v_i $(1 \leq i \leq n)$ is calculated as follows:

$$C_C(v_i) = \frac{n-1}{\sum_{j=1}^{n} d(v_i, v_j)} \tag{4.4}$$

where C_C denotes the closeness, $d(v_i, v_j)$ is the length of the shortest path from vertex v_i to vertex v_j, and n is the total number of vertices in the graph.

The measures of degree, betweenness and closeness describe the status of vertices. As degree only considers the relations between a vertex and its immediate neighbouring vertices, it is characterised as a local measure. In contrast, betweenness and closeness are considered as global measures, since they take the relations of a vertex and all other vertices into account. Specific to the objects in the real world, if an object has a high degree value, it means that the object directly connects to a large number of other objects, hence, it is a relatively important object. Betweenness evaluates to what extent a given object is part of the shortest paths that connect to any two other objects. A high betweenness value indicates that the object plays an important role in the connectivity of other objects as a ‘bridge’, hence, without this object, the connectivity of other objects might be broken. Closeness reflects how far (on average) an object is to every other object. This gives a sense to what extent an object is integrated or segregated with respect to other objects. A high closeness value denotes that the object is more integrated to all other objects, thus is more central and important.

4.2.2 Relative Trajectory Calculus (RTC)

The Relative Trajectory Calculus (RTC) was proposed by Van de Weghe (2004) as a qualitative approach to represent the spatio-temporal relationships between two disjoint moving objects based on describing their relative trajectories. Important in this calculus is that an object moving during a time interval I, which starts at timestamp t_1 and ends at timestamp t_2, is represented by means of a vector starting at t_1 and ending at t_2. Two objects moving during the same time interval can be represented by two vectors, each corresponding to a specific object. Using a single character label to denote the distance relation between the two objects, the RTC relationship between the two objects is determined. In the following, the RTC relationship is introduced in detail.

Assume: moving objects k and l, and timestamp t

$k|t$ denotes the position of k at t

$l|t$ denotes the position of l at t

$d(u, v)$ denotes the Euclidean distance between two positions u and v

$t_1 \prec t_2$ denotes that t_1 is temporally before t_2

$-$: the distance between k and l decreases:

$$\exists t_1, t_2 \left(t_1 \prec t \prec t_2 \land \forall t^-, t^+ \left(t_1 \prec t^- \prec t \prec t^+ \prec t_2 \rightarrow d(k|t^-, l|t^-) > d(k|t^+, l|t^+) \right) \right)$$

0: the distance between k and l remains the same:

$$\exists t_1, t_2 \left(t_1 \prec t \prec t_2 \land \forall t^-, t^+ \left(t_1 \prec t^- \prec t \prec t^+ \prec t_2 \rightarrow d(k|t^-, l|t^-) = d(k|t^+, l|t^+) \right) \right)$$

$+$: the distance between k and l increases:

$$\exists t_1, t_2 \left(t_1 \prec t \prec t_2 \land \forall t^-, t^+ \left(t_1 \prec t^- \prec t \prec t^+ \prec t_2 \rightarrow d(k|t^-, l|t^-) < d(k|t^+, l|t^+) \right) \right)$$

According to the relationship syntax between two moving objects, three RTC relations can be distinguished, as illustrated in Figure 4.1. For example, the RTC relation '–' indicates that the distance between the two objects before t is larger than the distance between the two objects after t. As the relations can represent the inter-relationship between two moving objects, they can be adopted to denote specific types of interaction patterns. Hence, three types of interaction patterns (i.e., attraction pattern, stability pattern and avoidance pattern) can be derived based on the RTC relations. The relationship between the RTC relations and the three types of interaction patterns is shown in Table 4.1.

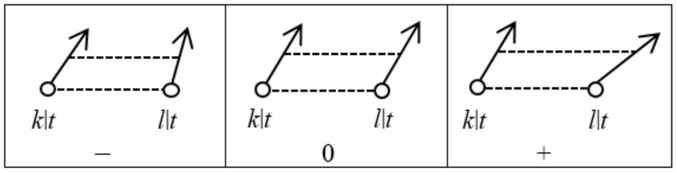


Figure 4.1. Illustration of the RTC relations.

Table 4.1. The relationship between the RTC relations and the three types of interaction patterns.

RTC relations	Types of interaction patterns
–	Attraction pattern
0	Stability pattern
+	Avoidance pattern

4.2.3 Continuous Triangular Model (CTM)

The Continuous Triangular Model (CTM) is an extension of the Triangular Model (TM), a 2D representation of time intervals that was initially introduced by Kulpa (1997). In the TM, any time interval is represented by a corresponding point. For example, in Figure 4.2, the time interval I, which starts at t_1 and ends at t_2, is represented by the intersection point P of two corresponding lines L_1 and L_2. In other words, point P equals to time interval I (or $[t_1, t_2]$). Hence, any sub time interval within I (or $[t_1, t_2]$) can be represented by a corresponding point within the triangle consisting of points P, t_1 and t_2. However, the TM appears incapable to represent the time intervals continuously. Hence, it was extended to the CTM by Qiang et al. (2014) to represent continuous temporal data. The CTM has the ability to represent the attribute values during all time intervals. The attribute value during a time interval is calculated using a specific algebra operator (such as mean, summation and maximum) based on the attributes at the finest sampled timestamps within the time interval. The attribute values at the timestamps between any two neighbouring sampled timestamps can be calculated using interpolation. Through colour-coding, the continuous field of the CTM is displayed as an image, in which each colour denotes the corresponding attribute value of a specific time interval. Hence, the colour at a specific point in the CTM denotes the attribute value during a corresponding time interval. For example, Figure 4.3 illustrates the values of the mean speed of a football player during the first five minutes of a game using the CTM. In Figure 4.3, the value of the mean speed during any time interval within [2, 4] minutes corresponds to a specific point within triangle A.

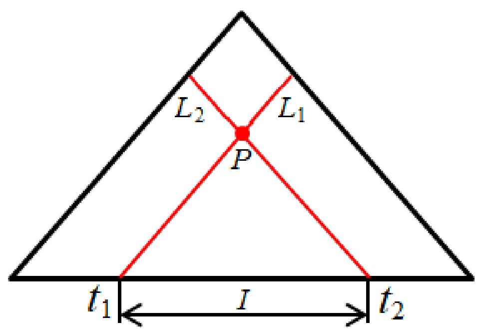


Figure 4.2. Illustration of the TM.

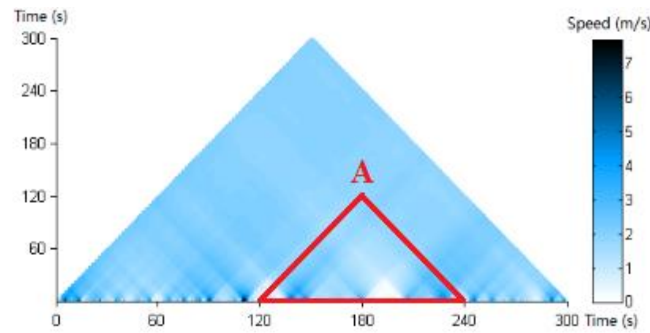


Figure 4.3. Illustration of the CTM.

In all, the CTM has a strong ability in visualising temporal information at any temporal scale, from the finest till the coarsest. Therefore, it can be employed as a multi-temporal scale tool, by which meaningful information, which cannot be revealed by other tools, might be discovered. Based on these characteristics, the CTM is utilised in this chapter to visualise the interaction patterns at multiple temporal scales so that useful information on dynamic interactions might be discovered (Zhang et al., 2016).

4.3 Methodology

In this chapter, we propose a hybrid approach combining the MTSSTN and the CTM to explore dynamic interactions in movement data. A MTSSTN is an extension of a network by taking space, time and temporal scale into account. The methodology mainly includes four steps: (1) the selection of a specific interaction pattern; (2) the generation of a MTSSTN; (3) the calculation of the interaction intensity measures and centrality measures, and (4) the visualisation and analysis of the results based on the CTM diagrams. In the following, these steps are introduced in detail.

4.3.1 The selection of a specific interaction pattern

The interaction patterns are derived based on RTC. As introduced in section 2, according to the RTC relations, three interaction patterns (i.e., attraction pattern, stability pattern and avoidance pattern) can be derived. Before generating MTSSTN, a specific interaction pattern has to be selected.

4.3.2 The generation of a MTSSTN

A MTSSTN is generated based on a corresponding interaction pattern. The generation of a MTSSTN includes three steps: (1) the generation of the vertices of the sub-networks during the time intervals at multiple temporal scales; (2) the generation of the edges of the sub-networks, and (3) the generation of the MTSSTN. A sample dataset, which is displayed in Figure 4.4(a), is utilised to illustrate how a MTSSTN is generated. In this dataset, three trajectories which are generated by three corresponding objects (i.e., O_1, O_2 and O_3) during five consecutive time intervals at the finest temporal scale are involved. In Figure 4.4(a), the (x, y) positions of each object at each of the six timestamps are given. Hence, the RTC relations of the three objects during the five consecutive time intervals at the finest temporal scale can be derived, which are shown in Figure 4.4(b).

4.3.2.1 The generation of the vertices of the sub-networks during the time intervals at multiple temporal scales

A sub-network is a part of a MTSSTN. More specifically, a sub-network is considered as a vertex of a MTSSTN. A sub-network is essentially a network during a time interval. A vertex in a sub-network denotes a corresponding moving object. For a moving object k, assume the spatial positions of k at timestamps t_0, t_1, t_2, …, t_n are (x_0, y_0), (x_1, y_1), (x_2, y_2), …, (x_n, y_n), respectively, for a time interval $[t_i, t_j]$ $(0 \leq i < j \leq n)$, the centroid of (x_i, y_i), (x_{i+1}, y_{i+1}), …, (x_q, y_q), …, (x_j, y_j) $(i < q < j)$ is considered as the position of the vertex corresponding to k during this time interval. The vertex is represented as $v_k|[t_i, t_j]$. As the CTM is an excellent tool to visually represent time intervals at multiple temporal scales, the sub-networks during time intervals at multiple temporal scales are visualised using the CTM-like format. According to this method, the generated vertices of the sub-networks during time intervals at multiple temporal scales are displayed in Figure 4.4(c). In Figure 4.4(c), the black points in each square denote the vertices of the sub-networks, and the squares denote the vertices of the MTSSTN. Besides, the temporal scale becomes coarser from the bottom to the top.

4.3.2.2 The generation of the edges of the sub-networks

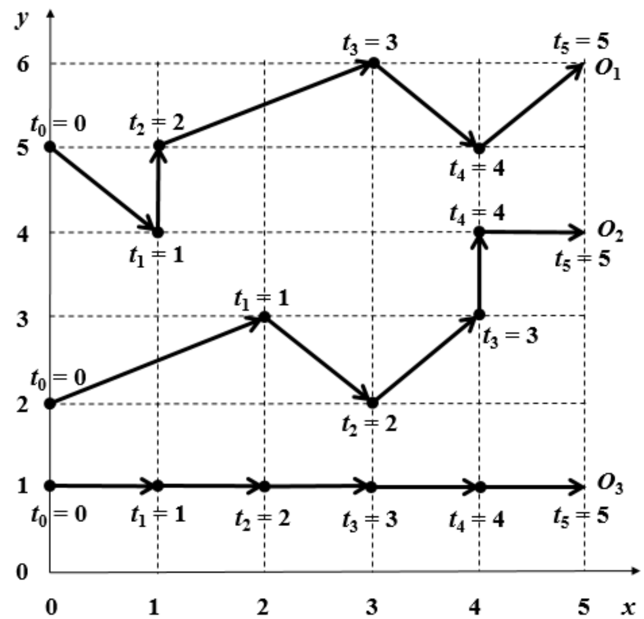
For an interaction pattern P, if two moving objects k and l form pattern P during a time interval [t_i, t_{i+1}] ($0 \leq i \leq n-1$) at the finest temporal scale, the corresponding value in the adjacency matrix is $r(v_k, v_l)|[t_i, t_{i+1}] = 1$, otherwise, $r(v_k, v_l)|[t_i, t_{i+1}] = 0$. For any time interval [t_i, t_j] ($0 \leq i < j \leq n$), the corresponding value in the adjacency matrix is the summation of that (i.e., the summation of the number of connections) during all the sub time intervals at the finest temporal scale. Hence, the corresponding value in the adjacency matrix is calculated as follows:

$$r(v_k, v_l)|[t_i, t_j] = \sum_{u=i}^{j-1} r(v_k, v_l)|[t_u, t_{u+1}] \tag{4.5}$$

If $r(v_k, v_l)|[t_i, t_j] \geq 1$, an edge exists between the two corresponding vertices $v_k|[t_i, t_j]$ and $v_l|[t_i, t_j]$, otherwise, not. Take the sample dataset and the attraction pattern for instance, the adjacency matrix during each of the time intervals at multiple temporal scales are calculated. The results are shown in Figure 4.4(d). Based on Figure 4.4(d), the edges of the sub-networks can be generated, which are shown in Figure 4.4(e).

4.3.2.3 The generation of the MTSSTN

A sub-network during a time interval is considered as a vertex of a MTSSTN. Hence, in this sub section, a vertex refers in particular to a sub-network mentioned above. For example, in Figure 4.4(e), each square is considered as a vertex of the final MTSSTN. When generating the edges of a MTSSTN, two principles are obeyed: (1) any two consecutive vertices at the same temporal scale are not allowed to be connected, and (2) any two vertices at two consecutive temporal scales are connected if the time intervals corresponding to the two vertices share a common time interval. According to these principles, the final MTSSTN can be generated. The generated MTSSTN corresponding to the sample dataset is shown in Figure 4.4(f).



(a)

Time interval: [0, 1]

RTC relation:

	O_1	O_2	O_3
O_1		−	−
O_2	−		+
O_3	−	+	

Time interval: [1, 2]

	O_1	O_2	O_3
O_1		+	+
O_2	+		−
O_3	+	−	

Time interval: [2, 3]

	O_1	O_2	O_3
O_1		−	+
O_2	−		+
O_3	+	+	

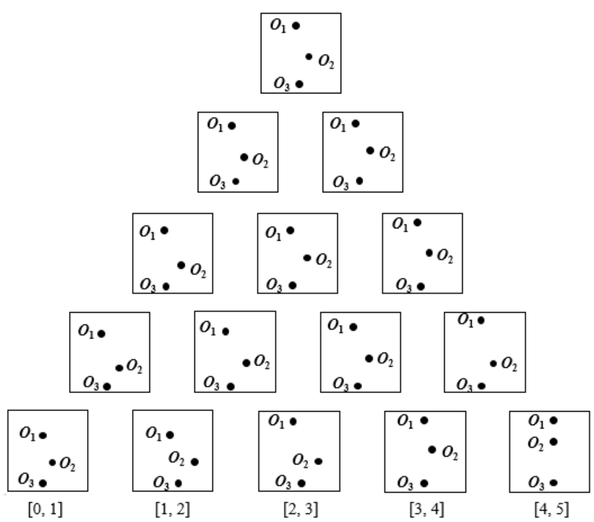
Time interval: [3, 4]

	O_1	O_2	O_3
O_1		−	−
O_2	−		+
O_3	−	+	

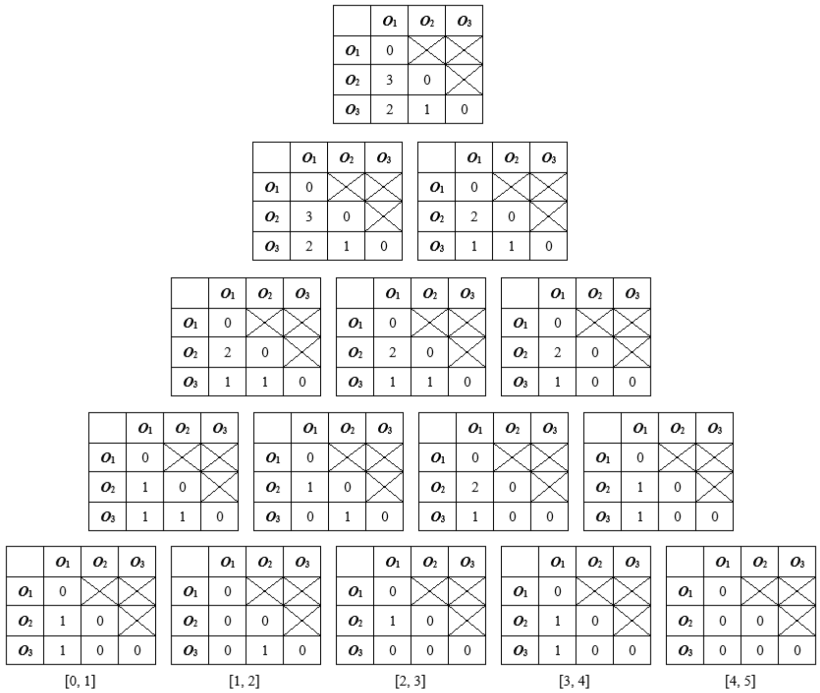
Time interval: [4, 5]

	O_1	O_2	O_3
O_1		+	+
O_2	+		0
O_3	+	0	

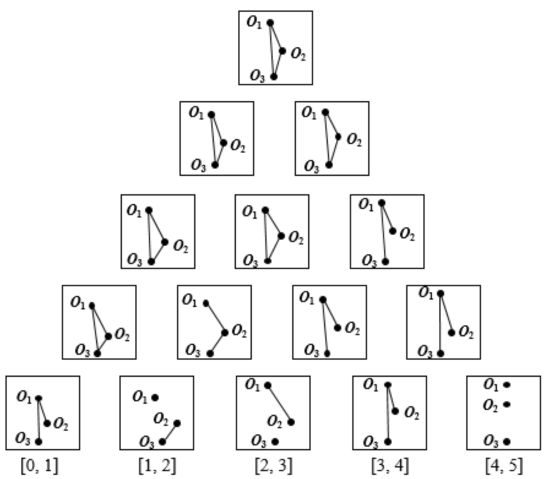
(b)



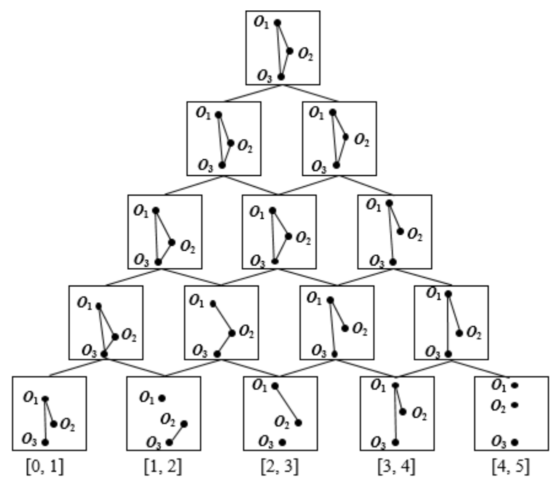
(c)



(d)



(e)



(f)

Figure 4.4. Illustration of the generation of a MTSSTN: (a) the three sample trajectories during five consecutive time intervals at the finest temporal scale, (b) the RTC relations during each time interval at the finest temporal scale, (c) the generated vertices of the sub-networks, (d) the adjacency matrix generated based on the attraction pattern during the time intervals at multiple temporal scales, (e) the generated edges of sub-networks, and (f) the generated MTSSTN.

4.3.3 The calculation of the interaction intensity measures and centrality measures

In this chapter, we define interaction intensity between two individuals during a time interval as the number of edges between the two corresponding vertices in the corresponding sub-networks at the finest temporal scale in the MTSSTN divided by the number of edges that they can have in theory. Hence, for two moving objects *k* and *l*, the interaction intensity between *k* and *l* in the sub-network during time interval [t_i, t_j] ($0 \leq i < j \leq n$) is calculated according to equation (4.6):

$$I_I(v_k, v_l)|[t_i, t_j] = \frac{\sum_{u=i}^{j-1} r(v_k, v_l)|[t_u, t_{u+1}]}{j - i} \tag{4.6}$$

where I_I denotes the interaction intensity, and *r* has the same meaning as in equation (4.5). The larger I_I is, the stronger the interaction between the two objects.

For at least three, the interactions among the individuals can be categorised as local interactions and global interactions. Local interactions denote the overall interaction between one individual and the others, while global interactions denote the overall interaction between all individuals. Assume *m* moving objects (i.e., O_1, O_2, O_3, …, O_m), the local interaction intensity measure and global interaction intensity measure can be calculated according to equations (4.7) and (4.8), respectively:

$$LI_I(v_{O_p})|[t_i, t_j] = \sum_{q=1}^{m} I_I(v_{O_p}, v_{O_q})|[t_i, t_j] \tag{4.7}$$

$$GI_I|[t_i, t_j] = \sum_{p=1}^{m-1} \sum_{q=p+1}^{m} I_I(v_{O_p}, v_{O_q})|[t_i, t_j] \tag{4.8}$$

where LI_I denotes the local interaction intensity of one individual, GI_I the global interaction intensity of all the individuals, and I_I the interaction intensity between two individuals.

The calculation of centrality measures in MTSSTN is similar to that in traditional networks. Specifically, the degree, betweenness and closeness of a moving object *k* during time interval [t_i, t_j] ($0 \leq i < j \leq n$) in the MTSSTN are calculated according to equations (4.9), (4.10) and (4.11), respectively:

$$C_D(v_k)|[t_i,t_j] = \sum_{u=1}^{n} r(v_k,v_u)|[t_i,t_j] \tag{4.9}$$

$$C_B(v_k)|[t_i,t_j] = \frac{\sum_{u=1}^{n-1}\sum_{v=k+1}^{n} p_{v_u v_k v_v}|[t_i,t_j]}{\sum_{u=1}^{n-1}\sum_{v=k+1}^{n} p_{v_u v_v}\,|[t_i,t_j]} \tag{4.10}$$

$$C_C(v_k)|[t_i,t_j] = \frac{n-1}{\sum_{u=1}^{n} d(v_k,v_u)|[t_i,t_j]} \tag{4.11}$$

where C_D, C_B and C_C denote degree, betweenness and closeness, respectively. The larger C_D/C_B/C_C is, the more important the object.

4.3.4 The visualisation and analyses of results based on the CTM diagrams

For each interaction pattern, both the interaction intensity measures and the centrality measures can be visualised using the CTM. Based on the corresponding CTM diagrams of the interaction intensity measures, the interaction intensities between any two individuals, or among multiple individuals can be explored. Based on the corresponding CTM diagrams of the centrality measures, the importance of the individuals in maintaining each interaction pattern can be explored. By applying the map algebra operator of ‘maximum’ to the CTM diagrams of the centrality measures of all individuals, the most important individuals in maintaining each interaction pattern can be identified.

4.4 Case study

4.4.1 Dataset

The movement data adopted in this chapter come from a real and entire football match between ‘Club Brugge KV’ and ‘Standard de Liège’ which took place on March 2nd 2014. For simplicity, we call them ‘Club Brugge’ and ‘Standard Liège’ respectively in the remainder of the chapter. Football is considered as a highly interactive sport since the players need to interact (e.g., collaborate) frequently with the teammates. As such, various types of interaction patterns can be involved during a match. The exploration of player interactions is important as player interactions can give insight into a team's playing style and can be used to assess the importance and performance of individual players of the team. In this dataset, the positions of all the players were tracked at a temporal resolution of 0.1s. The data include both spatio-temporal information and semantic information. The spatio-temporal information is recorded in a (*id*, *x*, *y*, *t*) format, where *id* identifies a specific player, *x* and *y* represent the *x* and *y* coordinates of the player's position, and *t* denotes the corresponding timestamp. The semantic information mainly includes the information of both teams, especially the basic information of the players (e.g., names, *id* numbers and positions played) and the events that happened during the match (e.g., event name, time of occurrence and id of the actors).

Note that Club Brugge won the match by 1-0 by scoring a goal at the timestamp 4733s. As relatively more sophisticated interactions might be involved in a relatively short time interval before a goal event, the movement data of the players (except the goalkeeper) of Club Brugge during the time interval before and until the goal event are used as the experiment data. After checking the semantic information in the original data, time interval [4493, 4733]s is used since it contains many interesting events (e.g., shot events and free-kick events) just before the goal was scored. Besides, in order to reduce the computational complexity, we down-sampled the temporal resolution from 0.1s to 1s. Note that for reasons of simplicity, the time interval is changed to [0, 240]s from [4493, 4733]s. In addition, due to privacy issues, the actual names of the 10 players are all replaced by player 1, player 2, player 3, …, and player 10.

4.4.2 Results and analysis

4.4.2.1 The interaction intensities between two individuals for each interaction pattern

Based on the proposed approach, the interaction intensities between any two individuals during all time intervals for each interaction pattern can be explored. Take for instance the attraction pattern, the interaction intensities between player 1 and player 6 are shown in

Figure 4.5, in which a darker colour denotes a stronger interaction and a lighter colour corresponds to a weaker interaction. For example, according to Figure 4.5, we can notice that the interactions between the two players during time interval just before the goal, [3.7, 3.8] minutes were relatively strong. In order to validate this, we plotted the trajectories of the two players during this time interval, which is shown in Figure 4.6, in which the red line denotes the trajectory of player 1, the green line the trajectory of player 6, and the black dotted line corresponds to the distance between the two players at each timestamp. Obviously, from Figure 4.6, we can clearly see that the distance between player 1 and player 6 decreased during each time interval from the 3.7^{th} minute to the 3.8^{th} minute. One is thus able to explore the interactions between any two individuals based on the corresponding CTM diagrams according to the specific demands (e.g., particularly interested in the movements of players during a specific time interval by watching the video), after which the performance of a pair of players can be explored, which might be important to sports professionals (e.g., coaches).

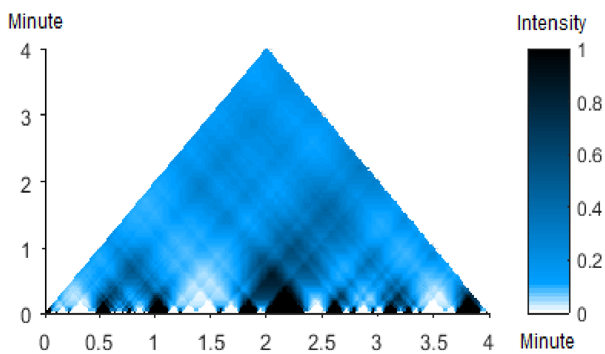


Figure 4.5. The interaction intensities between player 1 and player 6 during all time intervals for the attraction pattern.

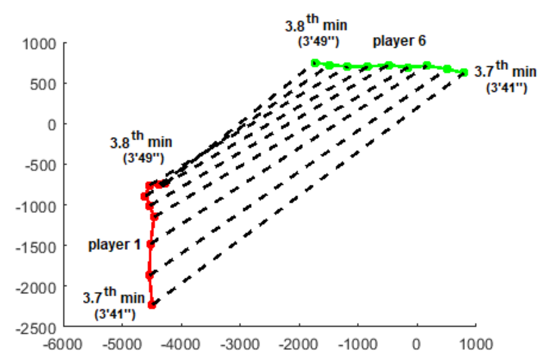


Figure 4.6. The trajectories of player 1 and player 6 during time interval [3.7, 3.8] minutes.

4.4.2.2 The interaction intensities among multiple individuals for each interaction pattern

Under many circumstances, one individual might interact with more than one other individual. Specifically in team sports, tactics may involve multiple players simultaneously. For example, as the most fundamental aspect of a football match, the passing of the ball always involves multiple players. Thus, the exploration of the overall interaction intensities between one player and other players (i.e., the local interaction intensity of one player) is important to related sports professionals. This can be achieved using the proposed approach. By checking the dataset, we select four players (i.e., player 1, player 2, player 5 and player 6) as an illustration, since they were involved in a sequence of passes before the goal. Take the attraction pattern for example, the local interaction intensities of each player are shown in Figure 4.7. From Figure 4.7, we can observe that each player had time intervals during which his general interactions with others were relatively strong. For instance, by comparing Figure 4.7(a), (b), (c) and (d), we can find that during the time interval around [3.7, 3.8] minutes, the local interactions of each of the four players were all relatively strong, which indicates that each of them may collaborate well with the others during this time interval. When looking at the football match, we can see that this interval of strong local interactions coincides with spatial compression of these four players just before the goal.

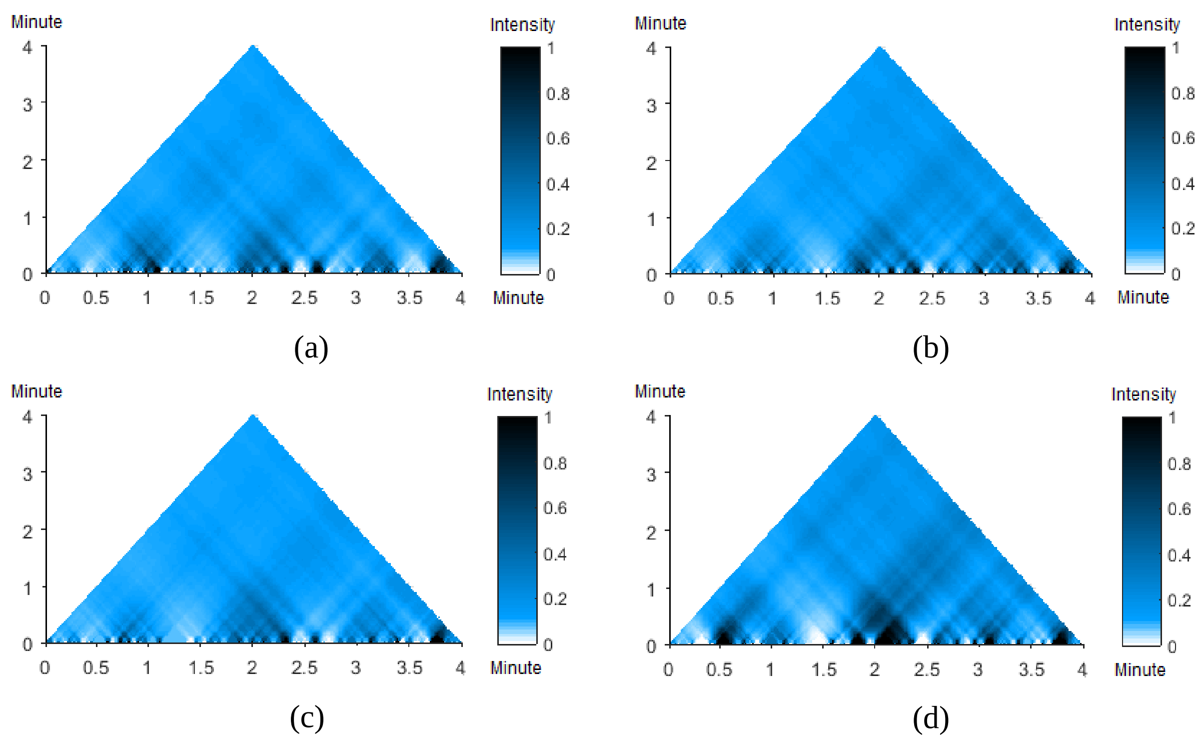


Figure 4.7. The local interaction intensities of each of the four players for the attraction pattern: (a) player 1, (b) player 2, (c) player 5, and (d) player 6.

Besides, as introduced in (Zhang et al., 2016), by employing corresponding map algebra operations, additional information can be discovered. In this case, a new CTM diagram is generated by employing the 'maximum' operator to the four CTM diagrams displayed in Figure 4.7. The new CTM diagram is shown in Figure 4.8, in which each colour corresponds with a specific player. Figure 4.8 depicts which player had the strongest local interactions during all time intervals. For instance, from Figure 4.8, we can conclude that player 6 interacted comparatively more intensively with the other three players from the perspective of a long time interval (e.g., longer than about 2.5 minutes). When the time intervals were shorter than about 2.5 minutes, each of the four players had specific time intervals during which he interacted comparatively more intensively with the other three players. This shows that, the proposed approach has potential to help sports professionals (e.g., coaches) to explore the local interactions of a group of target players in order to examine which player had a high influence on the movement patterns of the players around him, thus allowing evaluating player performance.

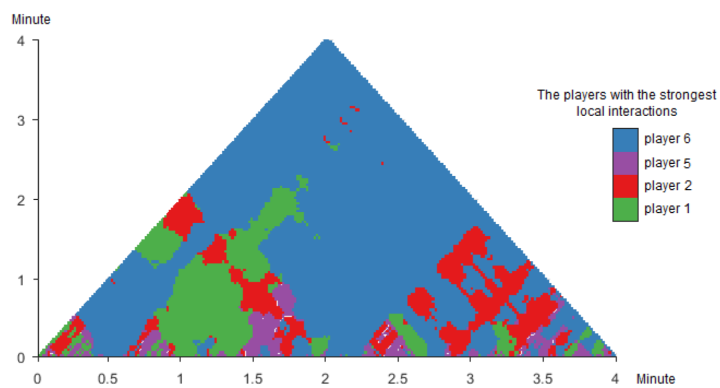


Figure 4.8. The players which had the strongest local interactions for the attraction pattern.

Apart from exploring the local interaction intensities of each individual for each interaction pattern, the approach can be used to explore the global interaction intensities of multiple individuals as well. This is particularly useful in team sports when examining the overall performance of multiple players. Take for instance all the ten players in a football match. When a team (e.g., team 1) possesses the ball, the players in the other team (e.g., team 2) tend to run towards each other to compress the space in order to tackle the opponents and to limit their options to pass the ball. Thus, for a good performance, the attraction pattern is expected to happen in team 2. On the contrary, the players of team 1 are expected to run away from each other in order to extend the space to pass the ball. Hence, for a good performance, the avoidance pattern is expected to happen in team 1. The overall performance of a team can be examined using the proposed approach by calculating the global interaction intensities among all the players. Take the avoidance pattern for instance, the global interaction intensities among all the players are shown in Figure 4.9.

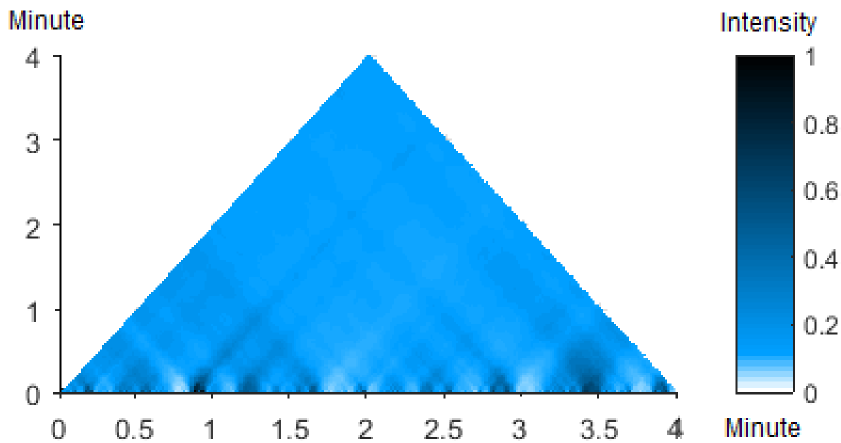


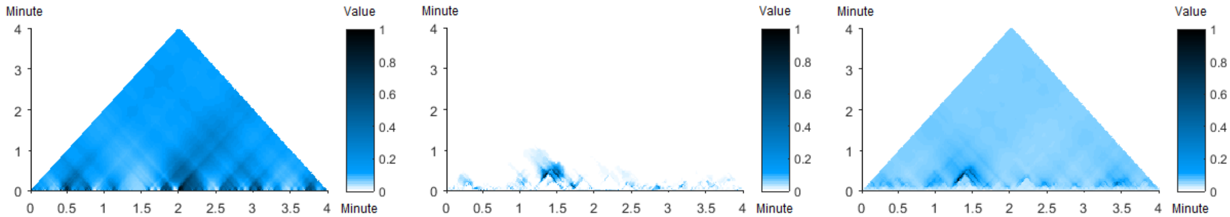
Figure 4.9. The global interaction intensities among all the players for the avoidance pattern.

In Figure 4.9, for example, we can notice that the global interaction intensity values during time interval [0.85, 0.95] minutes were relatively large (which indicates that the overall interactions during this time interval were relatively strong), and the global interaction intensity values during time interval [2.95, 3.10] minutes were relatively small (which indicates that the overall interactions during this time interval were relatively weak). By checking the adopted dataset, we find that during time interval [0.85, 0.95] minutes, Club Brugge possessed the ball, thus the avoidance pattern was expected to occur in case of a good performance. This coincides well with the results during the corresponding time interval in Figure 4.9, which demonstrates that Club Brugge indeed performed well generally during this time interval. On the other hand, during time interval [2.95, 3.10] minutes, Standard Liège possessed the ball. In case of a good performance, the attraction pattern was thus expected to happen for Club Brugge, which should make the overall interactions in the avoidance pattern weak during this time interval. Obviously, this coincides well with the results in Figure 4.9 as well. Based on this, we can infer that Club

Brugge also indeed had a good performance during this time interval. In terms of the simple analysis, we can conclude that the proposed approach has the ability to assist coaches to explore the overall interaction intensities among multiple players for each interaction pattern.

4.4.2.3 The importance of each individual and the identification of the most important individuals in each interaction pattern

The centrality measures can be used to measure the importance of an individual in an interaction pattern from different perspectives. In this case study, based on the centrality measures, the importance of each player for each interaction pattern can be evaluated, which might provide insightful information to coaches. Take player 6 and the attraction pattern for instance, the results of the centrality measures are visualised in Figure 4.10. In Figure 4.10, a dark colour corresponds to a high value, which means that the player was a central player during the corresponding time intervals according to the corresponding centrality measure. Note that 'central player' here and hereafter means a player playing an important role according to centrality measures, but not a player who is located at the spatial center of the field. In Figure 4.10, we can find that the importance of player 6 was different during all time intervals based on the same measure. For example, based on the measure of degree, player 6 can be considered as a central player during the time intervals which are dark, because player 6 had more direct connections with the other players in the MTSSTN during these time intervals. This indicates that the number of players whose distances between player 6 decreased during these time intervals was relatively large. Thus, player 6 had relatively good interactions with the other players for the attraction pattern and was considered a relatively central player of the movement pattern. Hence, based on the proposed approach, the importance of each player in each interaction pattern can be explored by using different centrality measures. This might provide potential insights to sports professionals and coaches for arranging suitable tactics and selecting the starting lineup for a match.



(a) (b) (c)

Figure 4.10. The visualisation of the centrality measures of player 6 for the attraction pattern: (a) degree, (b) betweenness, and (c) closeness.

The identification of important individuals is of high importance since they might play rather important roles in specific types of interactions. In this case study, the most central players for each interaction pattern during all time intervals can be identified based on the proposed approach. This is achieved by employing corresponding map algebra operators ('maximum' in this case) to the corresponding CTM diagrams of the centrality measures of each player. The detailed principle can be seen in (Zhang et al., 2016). The results are shown in Figure 4.11. Note that in Figure 4.11, similar colours are used for players at similar positions (i.e., players 1~3 are forward, players 4~6 midfielders and players 7~10 defenders). Figure 4.11 clearly demonstrates which player was the most central player during which time interval for each interaction pattern based on different centrality measures. For instance, for the attraction pattern (Figure 4.11(a)), player 4 can be considered as playing key roles based on degree and player 1 based on betweenness on a whole during relatively long time intervals (e.g., longer than about 1 minute). When the time intervals were shorter than 1 minute, each player had his own dominant time intervals, during which this player was considered as a central player. Similarly, for the stability pattern, players 1, 3 and 5 can be considered as the most central players on a whole based on degree, players 1 and 8 the most central in general based on betweenness, and players 1, 3, 5 and 6 the most central generally based on closeness. For the avoidance pattern, obviously, player 1 was the most central player on a whole based on betweenness. From Figure 4.11 we can observe that the results vary a lot, as in some of the figures, the most central players are quite easy to be identified, while in others this is not possible. However, one can draw specific conclusions by analysing the corresponding figures in depth (e.g., zooming in the zone of a specific time interval on a CTM diagram) according to specific demands. Therefore, this approach has the potential to provide insightful suggestions to related sports professionals for suitable tactical arrangement and players performance evaluation.

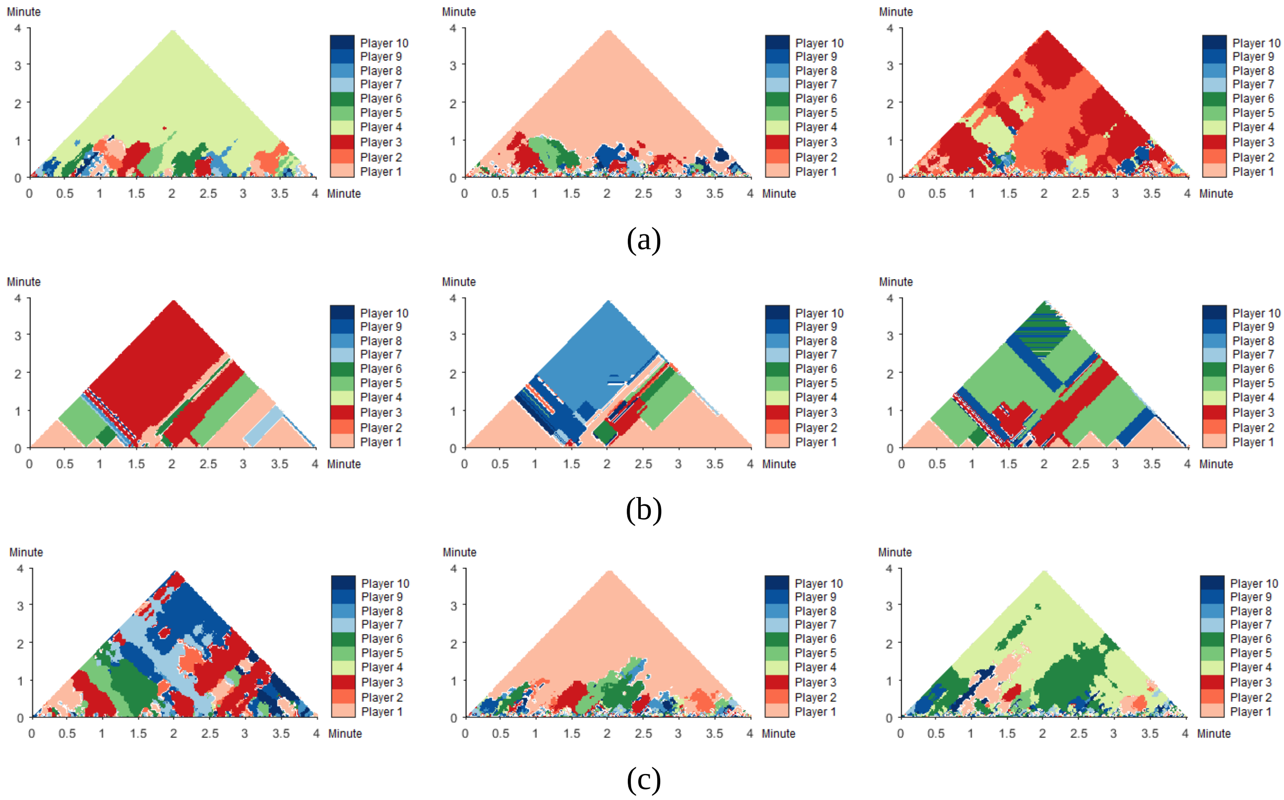


Figure 4.11. The most central players in each interaction pattern during all time intervals based on the centrality measures (from left to right: degree, betweenness and closeness): (a) the attraction pattern, (b) the stability pattern, and (c) the avoidance pattern.

4.5 Discussion

As a key contribution of this chapter, we develop a hybrid approach combining the Multi-Temporal Scale Spatio-Temporal Network and the Continuous Triangular Model and addressed the applicability of the approach in exploring dynamic interactions in movement data. Specifically, the approach is utilised in football movement data, a type of sports movement data which has gained much attention in recent years. The results show that the proposed approach can be used to explore various interactions between the players.

Besides, it is useful in evaluating the importance of each player and identifying the most central players.

As is known, scale is a common problematic issue in many disciplines, especially those that involve space and/or time (e.g., GIScience). In GIScience, scale is of significant importance. It even has been considered as the fifth dimension in 5D data modelling (van Oosterom & Stoter, 2010). Thus, it is crucial to take scale into consideration when dealing with space and/or time related data. However, only limited research on dynamic interactions in movement data has taken scale (or in detail, temporal scale) into consideration. The research conducted by Long & Nelson (2013) can be regarded as a typical example. In this research, it was argued that dynamic interactions can be analysed from four analysis levels, i.e., local, interval, episodal and global. Local level is described as the finest temporal scale, interval level and episodal level correspond to a coarser temporal scale, and global level is considered as the coarsest temporal scale. A new computation has to be made if the analysis level changes, and new results have to be visualised correspondingly. Actually, the results at the four levels are contained simultaneously in the results of the proposed approach in this chapter. In other words, the results at any of the four levels can be found in one corresponding CTM diagram.

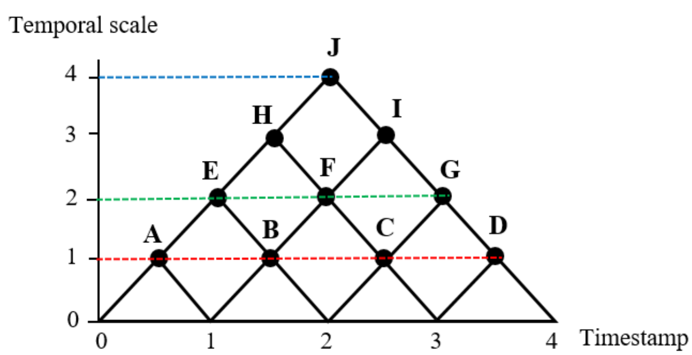


Figure 4.12. Illustration of the local level (red dotted line), interval level (green dotted line), episodal level (green dotted line) and global level (blue dotted line) in the CTM.

Take the simple CTM diagram in Figure 4.12 for example, the *x*-coordinate and *y*-coordinate denote the timestamps and the temporal scales, respectively, the red dotted line and the blue dotted line indicate the local level and the global level, respectively, and the green dotted line either an interval level or an episodal level. Therefore, the values at points A, B, C and D correspond to the values at the local level, and the value at point J corresponds to the value at the global level. The values at the interval level can be found at points E, F and G, and the values at the episodal level can be found at points E and G. Hence, this approach appears to be superior in analysing dynamic interactions at multiple temporal scales.

The dynamic interactions in movement data include both the interactions between a pair of individuals and the interactions among multiple (usually at least three) individuals. So far, current methods have mainly focused on the interactions between two individuals. However, it is common in practice to consider several objects as a group (e.g., four players as a moving flock). Hence, a method to explore the interactions among more than two individuals simultaneously is very much in need. Benefitting from the map algebra operations supported by the CTM, the proposed approach is capable to explore the interactions among multiple individuals. The results show the effectiveness of this approach.

Network science provides a lot of powerful methods to study systems in the real world. Therefore, networks have been widely adopted as a useful tool to enable researchers to explore many systems in society, nature and technology. In the area of movement data analysis, networks have already been utilised as a tool as well. In recent time, a new type of network (i.e., Spatio-Temporal Network) has been developed by Williams & Musolesi (2016) in order to analyse spatio-temporal data more accurately. However, it appears unsatisfactory to analyse multi-temporal scale related issues. Based on this, in this chapter, we propose an even more novel type of network (i.e., Multi-Temporal Scale Spatio-Temporal Network). To the best of our knowledge, this is the first case that adopted networks as a tool to analyse spatio-temporal data at multiple temporal scales. The results reveal the advantages of using the Multi-Temporal Scale Spatio-Temporal Network to discover insightful information that current networks cannot. We expect that the new type of network could be considered as a potential tool to gradually serve the domain of data analysis in the future.

However, the proposed approach still has its own flaws. On the one hand, although the interaction patterns derived based on RTC are prevalent between/among any two/multiple objects, they are still relatively simplistic. This, to some extent, limits the significance of the results exhibited during all time intervals in the case study. The results primarily make sense quite well during part of the time intervals, which are chosen by users (e.g., coaches) based on relevant information (e.g., the video of the match) according to their demands. In the future, more sophisticated interaction patterns might be derived using other potential methods automatically, such as Qualitative Trajectory Calculus (QTC) (Van de Weghe et al., 2004; Van de Weghe et al., 2005; Van de Weghe et al., 2006), Dynamic Interaction (DI) (Long & Nelson, 2013) and RElative MOtion (REMO) (Laube et al., 2005), in order to enhance the meanings of results. On the other hand, the current approach appears costly in time (the complexity is $O(n^3)$, where n is the number of time intervals at the finest temporal scale), thus limiting its applicability for very large movement datasets, although it showed its applicability and usefulness in relatively small datasets (e.g., it took approximately 4 minutes in this case study on a Windows 10 system with a processor of 2.6 GHz and a RAM of 8 GB). Things thus need to be optimised in the future.

4.6 Conclusions and future work

Currently, movement data are collected in a variety of domains and are becoming a popular type of data. Although many research topics with respect to movement data have been undertaken, the research on dynamic interactions in movement data is still in its infancy. This chapter proposes a hybrid approach combining the Multi-Temporal Scale Spatio-Temporal Network and the Continuous Triangular Model to explore dynamic interactions in movement data. Four main steps are included in the approach: first, RTC is used to derive three interesting interaction patterns; second, a corresponding MTSSTN is generated for each interaction pattern; third, for each MTSSTN, both the interaction intensity measures and the three centrality measures - degree, betweenness and closeness -

are computed; finally, the results are visualised using the CTM and analysed based on the CTM diagrams. Based on the proposed approach, three distinctive aims regarding the dynamic interactions in movement data can be achieved at multiple temporal scales. The approach is then validated based on the movement data obtained from a real football match. The results demonstrate that the dynamic interactions involved in movement data can be explored effectively and useful information can be discovered based on the proposed approach.

In this chapter, central to the approach is the construction of the MTSSTNs. For each interaction pattern, a corresponding MTSSTN can be generated. Networks include undirected or directed networks, and unweighted or weighted networks. As our aim is primarily methodological, in this chapter, only unweighted undirected networks are generated. In the future, directed networks can be generated so that more insightful information might be revealed. Besides, in the proposed approach, the MTSSTN is generated only based on the spatio-temporal information involved in the movement data. In the future, the semantic information can also be employed to generate new types of MTSSTNs. For example, based on the passing information in football movement data, the multi-temporal scale spatio-temporal passing networks can be derived, according to which the dynamic interactions of football players regarding passing can be explored. In addition, the movement data from other domains can also be employed in order to extend the range of applications of the proposed approach.

References

Ahearn, S. C., Dodge, S., Simcharoen, A., Xavier, G., & Smith, J. L. (2017). A context-sensitive correlated random walk: a new simulation model for movement. *International Journal of Geographical Information Science, 31*(5), 867-883.

Andrienko, G., Andrienko, N., Burch, M., & Weiskopf, D. (2012). Visual analytics methodology for eye movement studies. *IEEE Transactions on Visualization and Computer Graphics*, *18*(12), 2889-2898.

Andrienko, G., Andrienko, N., Demsar, U., Dransch, D., Dykes, J., Fabrikant, S. I., ... & Tominski, C. (2010). Space, time and visual analytics. *International Journal of Geographical Information Science*, *24*(10), 1577-1600.

Andrienko, G., Andrienko, N., & Wrobel, S. (2007). Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter*, *9*(2), 38-46.

Barrat, A., Barthélemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, *101*(11), 3747-3752.

Civilis, A., Jensen, C. S., & Pakalnis, S. (2005). Techniques for efficient road-network-based tracking of moving objects. *IEEE Transactions on Knowledge and Data Engineering*, *17*(5), 698-712.

Delafontaine, M., Versichele, M., Neutens, T., & Van de Weghe, N. (2012). Analysing spatiotemporal sequences in Bluetooth tracking data. *Applied Geography*, *34*, 659-668.

Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Van de Weghe, N., ... & Weibel, R. (2015). Analysis and visualisation of movement: an interdisciplinary review. *Movement ecology*, *3*(1), 5.

Doncaster, C. P. (1990). Non-parametric estimates of interaction from radio-tracking data. *Journal of Theoretical Biology*, *143*(4), 431-443.

Gomez, G., López, P. H., Link, D., & Eskofier, B. (2014). Tracking of ball and players in beach volleyball videos. *PloS one*, *9*(11), e111730.

Gudmundsson, J., & Wolle, T. (2014). Football analysis using spatio-temporal tools. *Computers, Environment and Urban Systems*, *47*, 16-27.

He, J., & Chen, D. (2015). A fast algorithm for community detection in temporal network. *Physica A: Statistical Mechanics and its Applications*, *429*, 87-94.

Holme, P. (2015). Modern temporal network theory: a colloquium. *The European Physical Journal B*, *88*(9), 234.

Hornsby, K., & Egenhofer, M. J. (2002). Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence*, *36*(1-2), 177-194.

Jiang, B., & Claramunt, C. (2004). A structural approach to the model generalization of an urban street network. *GeoInformatica*, *8*(2), 157-171.

Konzack, M., McKetterick, T., Ophelders, T., Buchin, M., Giuggioli, L., Long, J., ... & Buchin, K. (2017). Visual analytics of delays and interaction in movement data. *International Journal of Geographical Information Science*, *31*(2), 320-345.

Kulpa, Z. (1997). Diagrammatic representation of interval space in proving theorems about interval relations. *Reliable Computing*, *3*(3), 209-217.

Kveladze, I., Kraak, M. J., & Van Elzakker, C. P. (2015). The space-time cube as part of a GeoVisual analytics environment to support the understanding of movement data. *International journal of geographical information science*, *29*(11), 2001-2016.

Laube, P., Imfeld, S., & Weibel, R. (2005). Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science*, *19*(6), 639-668.

Lee, J. G., Han, J., & Whang, K. (2007). Trajectory clustering, a partition-and-group framework. In *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data* (pp. 593-604).

Lee, S., Rocha, L. E., Liljeros, F., & Holme, P. (2012). Exploiting temporal network structures of human interaction to effectively immunize populations. *PloS one*, *7*(5), e36439.

Long, J. A. (2015). Quantifying spatial-temporal interactions from wildlife tracking data: issues of space, time, and statistical significance. *Procedia Environmental Sciences*, *26*, 3-10.

Long, J. A., Nelson, T. A., Webb, S. L., & Gee, K. L. (2014). A critical examination of indices of dynamic interaction for wildlife telemetry studies. *Journal of Animal Ecology*, *83*(5), 1216-1233.

Long, J. A., & Nelson, T. A. (2013). Measuring dynamic interaction in movement data. *Transactions in GIS, 17*(1), 62-77.

Meijles, E. W., De Bakker, M., Groote, P. D., & Barske, R. (2014). Analysing hiker movement patterns using GPS data: Implications for park management. *Computers*, *Environment and Urban Systems*, *47*, 44-57.

Miller, J. A. (2012). Using spatially explicit simulated data to analyze animal interactions: a case study with brown hyenas in Northern Botswana. *Transactions in GIS*, *16*(3), 271-291.

Miller, J. A. (2015). Towards a better understanding of dynamic interaction metrics for wildlife: a null model approach. *Transactions in GIS*, *19*(3), 342-361.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, *45*(2), 167-256.

Qiang, Y., Chavoshi, S. H., Logghe, S., De Maeyer, P., & Van de Weghe, N. (2014). Multi-scale analysis of linear data in a two-dimensional space. *Information Visualization*, *13*(3), 248-265.

Shamoun-Baranes, J., van Loon, E. E., Purves, R. S., Speckmann, B., Weiskopf, D., & Camphuysen, C. J. (2012). Analysis and visualization of animal movement. *Biology Letters*, *8*(1), 6-9.

Van de Weghe, N. (2004). *Representing and reasoning about moving objects: a qualitative approach*. Ghent University.

Van de Weghe, N., Cohn, A. G., De Maeyer, P., & Witlox, F. (2005). Representing moving objects in computer-based expert systems: the overtake event example. *Expert Systems with Applications*, *29*(4), 977-983.

Van de Weghe, N., Cohn, A. G., De Tre, G., & De Maeyer, P. (2006). A qualitative trajectory calculus as a basis for representing moving objects in geographical information systems. *Control and Cybernetics*, *35*(1), 97-119.

Van de Weghe, N., Maddens, R., Bogaert, P., Brondeel, M., & De Maeyer, P. (2004). Qualitative analysis of polygon shape-change. In *IEEE International Geoscience and Remote Sensing Symposium* (pp. 4157-4159).

van Oosterom, P., & Stoter, J. (2010). 5D data modelling: full integration of 2D/3D space, time and scale dimensions. *Lecture Notes in Computer Science*, *6292*, 310-324.

Vogel, M., Hamon, R., Lozenguez, G., Merchez, L., Abry, P., Barnier, J., ... & Robardet, C. (2014). From bicycle sharing system movements to users: a typology of Vélo’ cyclists in Lyon based on large-scale behavioural dataset. *Journal of Transport Geography*, *41*, 280-291.

Von Landesberger, T., Brodkorb, F., Roskosch, P., Andrienko, N., Andrienko, G., & Kerren, A. (2016). Mobilitygraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering. *IEEE transactions on visualization and computer graphics*, *22*(1), 11-20.

Wang, J., Duckham, M., & Worboys, M. (2016). A framework for models of movement in geographic space. *International Journal of Geographical Information Science*, *30*(5), 970-992.

Wang, Y., Luo, Z., Takekawa, J., Prosser, D., Xiong, Y., Newman, S., ... & Yan, B. (2016). A new method for discovering behavior patterns among animal movements. *International Journal of Geographical Information Science*, *30*(5), 929-947.

Williams, M. J., & Musolesi, M. (2016). Spatio-temporal networks: reachability, centrality and robustness. *Royal Society open science*, *3*(6), 160196.

Xu, Z., Sandrasegaran, K., Kong, X., Zhu, X., Zhao, J., Hu, B., & Lin, C. C. (2013). Pedestrian monitoring system using Wi-Fi technology and RSSI based localization. *International Journal of Wireless & Mobile Networks*, *5*(4), 17-34.

Zhang, P., Beernaerts, J., Zhang, L., & Van de Weghe, N. (2016). Visual exploration of match performance based on football movement data using the Continuous Triangular Model. *Applied Geography*, *76*, 1-13.

Zhang, P., Deng, M., Shi, Y., & Zhao, L. (2017). Detecting hotspots of urban residents' behaviours based on spatio-temporal clustering techniques. *GeoJournal*, *82*(5), 923-935.

5

Discovering Moving Flock Patterns in Movement Data: A Reeb Graph-Based Approach

Modified from: Zhang, P., & Van de Weghe, N. (2018). Discovering moving flock patterns in movement data: a Reeb graph-based approach. (To be submitted)

Abstract: With the rapid development of location-aware technologies, a proliferation of rich and voluminous movement data have been resulted in. This thereby necessitates the various research topics with respect to movement data, such as the analysis, mining and visualisation of movement data. Among the many topics, the mining of movement patterns in movement data takes a large proportion. In this chapter, we mainly aim at discovering one of the typical movement patterns, i.e., moving flock patterns. Although a large body of research with respect to moving flock patterns has been undertaken, specific shortages still exist. Based on this, we first develop an improved definition of moving flock on the basis of the existing one. Then, a taxonomy of moving flock patterns is proposed, according to which eight types of interesting moving flock patterns are derived. Subsequently, we propose a Reeb graph-based approach for discovering moving flock patterns in movement data, and further use this approach to distinguish the eight types of moving flock patterns. The novel sports-oriented movement data, which were obtained from a real football match, are adopted as a case study to validate the effectiveness of the proposed approach and expand the application fields of the research of moving flock patterns. The results

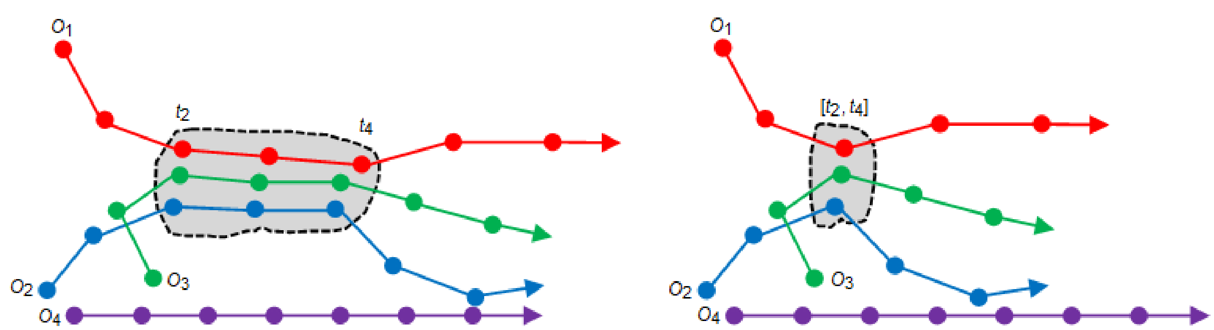
demonstrate that the proposed approach is capable of discovering the desired moving flock patterns and insightful information can be provided by applying the proposed approach to specific fields.

5.1 Introduction

Benefiting from the developments of location-aware technologies such as GPS (global positioning system), Bluetooth, RFID (radio frequency identification), WiFi, and image recgoniton, data related to the trajectories of moving objects can be acquired more easily than ever before. This has caused a proliferation of rich and voluminous movement data. Nowadays, a large variety of movement data are in use or already attracted attention. Specific types of movement data include transportation related data (Civilis et al., 2005; Delafontaine et al., 2012; Zhang et al., 2017), animal movement data (Laube et al., 2005; Shamoun-Baranes et al., 2012; Demšar et al., 2015), natural phenomena movement data (Lee et al., 2007), as well as sports movement data (Gudmundsson et al., 2014; Gomez et al., 2014; Zhang et al., 2016; Zhang et al., 2018). Given the emergence of the large amount of tracked movement data, the analysis of movement data has been paid more and more attention by researchers from various disciplines and fields, including geographical information science (GIScience), data mining and knowledge discovery, computational geometry, and so forth.

Among the current research, the mining of movement patterns takes a large proportion. The mining of movement patterns appears important because movement patterns can exhibit the rules of objects’ movements, which may imply important meanings in practice. However, in most cases, movement patterns are hidden behind the large amount of movement data. Thus, the approaches for mining movement patterns in movement data are much in need. Compared to the mining of movement patterns of individual moving objects, the mining of movement patterns of multiple moving objects has been attracting much more attention. Typical movement patterns existing in multiple moving objects include flock patterns (Laube & Imfeld, 2002; Laube et al., 2004; Laube et al., 2005; Gudmundsson

& van Kreveld, 2006; Gudmundsson et al., 2007; Benkert et al., 2008; Vieira et al., 2009; Wachowicz et al., 2011; Jacob & Idicula, 2012; Kjærgaard et al., 2012; Fort et al., 2014; Turdukulov et al., 2014; Cao et al., 2016), convoy patterns (Jeung et al., 2008; Yeoman & Duckham, 2016), leadership patterns (Laube et al., 2005; Gudmundsson et al., 2007; Andersson et al., 2008; Solera et al., 2015), moving clusters (Kalnis et al., 2005) and crews (Loglisci 2017). Among the listed movement patterns, the research on flock patterns takes a large proportion. Flocks are usually associated with the movements of a group of moving objects, such as birds, animals, pedestrians and vehicles. It can be informally depicted as 'a group of spatially close objects staying together for a specific time interval', as illustrated in Figure 5.1. Figure 5.1 illustrates the movements of four objects (i.e., O_1, O_2, O_3 and O_4) during seven consecutive timestamps. In Figure 5.1, as objects O_1, O_2 and O_3 stay closely together during time interval $[t_2, t_4]$, they can be considered as a flock pattern during this time interval. Flock patterns include moving flock patterns and stationary flock patterns. If all members involved in a flock pattern keep moving during the corresponding time interval, they can be regarded as a moving flock pattern. On the contrary, if all members keep stationary during the corresponding time interval, they are considered as a stationary flock pattern. Take Figure 5.1 for instance, in Figure 5.1(a), objects O_1, O_2 and O_3 form a moving flock pattern during time interval $[t_2, t_4]$ since they keep moving during this period. In Figure 5.1(b), objects O_1, O_2 and O_3 form a stationary flock pattern during time interval $[t_2, t_4]$ as they keep stationary during this period.



(a) (b)

Figure 5.1. Illustration of flock patterns: (a) moving flock pattern, and (b) stationary flock pattern (O_1, O_2, and O_3 are stationary during time interval $[t_2, t_4]$ while O_4 is moving).

The pioneering work on flock pattern discovery stems from (Laube & Imfeld, 2002; Laube et al., 2004; Laube et al., 2005). In their work, they proposed the concept ‘flock’ for the first time and defined flock as a set of objects that stay spatially close and exhibit similar properties in terms of motion attributes (e.g., speed, change of speed and motion azimuth) at some timestamp. They then designed corresponding algorithms based on the REMO (RElative MOtion) concept to discover flock patterns. The flock patterns that last for one timestamp indeed can be discovered using these algorithms. However, since they didn’t take a minimum lasting time into consideration, the flock patterns that last for more than one timestamp were not able to be discovered. Given this limitation, Benkert et al. (2008) extended the definition of flock to a new one, in which the minimum lasting time is taken into account. Specifically, they emphasised that the objects should stay together for at least a period of time, for instance, k ($k > 1$) timestamps, rather than just one single timestamp. Then, they simplified the definition by transferring flock patterns to approximate flock patterns and developed corresponding algorithms to discover approximate flock patterns based on skip-quadtree (Eppstein et al., 2005). Apart from the approximate flock patterns, Gudmundsson & van Kreveld (2006) proposed two other types of flock patterns: fixed-flock pattern and varying-flock pattern. Fixed-flock pattern means that the members of a flock pattern keep the same during the entire time interval while the members in a varying-flock pattern have changes during the entire time interval. They then generated corresponding algorithms to compute longest duration flock patterns from trajectory data. Besides, other researchers have also focused on developing various algorithms to discover flock patterns. Vieira et al. (2009) for the first time exhibited an on-line flock patterns discovery approach from spatio-temporal data and designed five corresponding algorithms. Wachowicz et al. (2011) proposed the first formal definition of moving flock and used collective coherence (Wood & Galton, 2009) to refer to the spatial closeness over some time interval with a minimum number of members. Then, they used it to find moving flock patterns among pedestrians. Kjærgaard et al. (2012) conducted the work on finding pedestrian flocks in indoor environments from the dataset collected via WiFi signals of mobile phones. Turdukulov et al. (2014) presented a frequent pattern mining approach for discovering moving flock patterns in large spatio-temporal datasets. The approach was combined with the visual interface of the space-time cube, which allows users to interactively explore the results. Fort et al. (2014) developed an efficient parallel GPU-based algorithm for reporting three different variants of flock patterns: all maximum flocks, the largest flock and the longest flock. The algorithm was tested on three real trajectory datasets and showed its efficiency and scalability. Cao et al. (2016) proposed a definition of 'freedom moving flock pattern' and a corresponding algorithm for extracting such pattern. The algorithm was tested on two real pedestrian trajectory datasets and the results showed the usefulness of extracting such pattern.

By summarising the aforementioned research, we can find that although different researchers proposed different types/classifications of movement patterns (e.g., a taxonomy of movement patterns presented in Dodge et al. (2008), there still lacks a general taxonomy of flock patterns and moving flock patterns. Besides, although Wachowicz et al. (2011) proposed the first formal definition of moving flock based on the definition of flock previously proposed by Benkert et al. (2008), there still exists specific drawbacks, as it might classify some stationary flock patterns wrongly as moving flock patterns. Based on this, in this chapter, we first develop an improved definition of moving flock so as to find more exact moving flock patterns. Second, a taxonomy of moving flock patterns is proposed, by which various types of moving flock patterns can be derived. Among the generated moving flock patterns, we then extract eight different types which appear quite interesting from our perspective and consider them as the main focus in this chapter. Finally, a Reeb graph-based approach is proposed in order to discover moving flock patterns and further the eight types of moving flock patterns.

The remainder of this chapter is organised as follows. Section 5.2 introduces the existing definition of moving flock, the improved definition of moving flock proposed by us, and the taxonomy of moving flock patterns. In section 5.3, the methodology for discovering moving flock patterns and further the eight types of moving flock patterns is described in detail. In section 5.4, a case study is conducted by using the movement data acquired from a real football match to validate the effectiveness of the proposed approach. Finally, section 5.5 gives the conclusions and some recommendations for future work.

5.2 Definition of moving flock and taxonomy of moving flock patterns

In this section, first, the definitions of moving flock, including both the previous one and the improved one, are introduced. Then, a taxonomy of moving flock patterns is proposed, according to which various types of moving flock patterns can be derived.

5.2.1 Definitions of flock and moving flock

There have already been several definitions of flock, such as the ones proposed by Laube et al. (2005), Gudmundsson & van Kreveld (2006), Benkert et al. (2008) and Vieira et al. (2009). Among them, the one defined by Benkert et al. (2008) is adopted in this chapter, as it is used frequently in existing research. It can be defined as follows.

Definition 5.1 (Flock): Given a set of *n* trajectories, an (*r*, *m*, *k*)-flock *F* during a time interval $I = [t_i, t_j]$, where $j - i \geq k$, consists of at least *m* objects such that for every discrete timestamp $t_l \in I$ $(i \leq l \leq j)$, there is a disk of radius *r* that contains all the *m* objects.

Nevertheless, this definition cannot distinguish between stationary flock and moving flock. Hence, Wachowicz et al. (2011) proposed the definition of moving flock based on Definition 5.1. To the best of our knowledge, this was the first formal definition of moving flock. The detailed definitions are defined below.

Definition 5.2 (Spatial extent of a flock): Given an (*r*, *m*, *k*)-flock *F* during a time interval *I*, its spatial extent, $ext(F, I)$, is defined as $ext(F, I) = max\{l, w\}$, where l and w are the length and width of the minimum bounding rectangle (MBR) of the set of sub trajectories belonging to the flock, respectively.

Definition 5.3 (Moving flock): Given a set of *n* trajectories, an (*r*, *m*, *k*)-moving flock F_M during a time interval $I = [t_i, t_j]$, where $j - i \geq k$, consists of at least *m* objects such that for every discrete timestamp $t_l \in I$ $(i \leq l \leq j)$, there is a disk of radius *r* that contains all the *m* objects and the spatial extent of F_M meets the requirement $ext(F_M, I) \geq r$.

However, it still appears deficient as a stationary flock might be misunderstood as a moving flock according to the definition. Take a collection of *m* objects for example, assume they were a flock during time interval $I = [t_i, t_j]$ $(j - i \geq k)$, but they kept stationary during $[t_i, t_{j-1}]$ and kept moving during $[t_{j-1}, t_j]$ (met the requirement $ext(F_M, I) \geq r$ when moving). According to Definition 5.2 and Definition 5.3, the *m* objects are considered as a moving flock during $[t_i, t_j]$. However, strictly speaking, they were a moving flock only during $[t_{j-1}, t_j]$ instead of the whole interval $[t_i, t_j]$. Hence, in this chapter, an improved definition of moving flock is proposed on the basis of the aforementioned definitions. The improved definition is described in detail as follows.

Definition 5.4 (Spatial extent of a flock between two consecutive timestamps): Given an (*r*, *m*, *k*)-flock *F* during a time interval $I = [t_i, t_j]$, where $j - i \geq k$, assume t_u and t_{u+1} $(i \leq u \leq j - 1)$ are two consecutive timestamps, the spatial extent of *F* between t_u and t_{u+1} is defined as $ext(F|t_u, F|t_{u+1}) = d\{p_u, p_{u+1}\}$, where p_u and p_{u+1} are the geometric centres of the MBRs of *F* at t_u and t_{u+1}, respectively, and $d\{p_u, p_{u+1}\}$ is the Euclidean distance between p_u and p_{u+1}.

Take Figure 5.2 for example, $ext(F|t_u, F|t_{u+1})$ denotes the spatial extent of a flock *F* (consisting of three objects, i.e., O_1, O_2 and O_3) between two consecutive timestamps t_u and t_{u+1}.

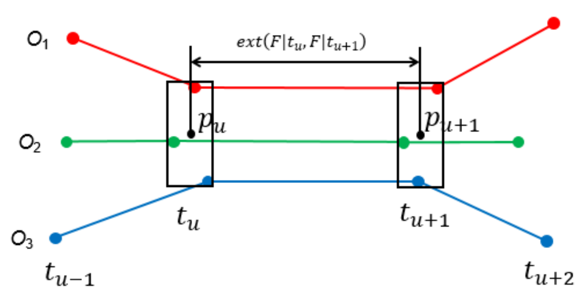


Figure 5.2. Illustration of the spatial extent of a flock between two consecutive timestamps.

Definition 5.5 (Moving flock): Given a set of *n* trajectories, an (*r*, *m*, *k*, *θ*)-moving flock

F_M in a time interval $I = [t_i, t_j]$, where $j - i \geq k$, consists of at least m objects such that for every discrete timestamp $t_l \in I$ $(i \leq l \leq j)$, there is a disk of radius r that contains all the m objects and any spatial extent between two consecutive timestamps $ext(F_M|t_u, F_M|t_{u+1}) \geq d$ $(i \leq u \leq j-1, d > 0)$. Assume the m objects are O_1, O_2, …, O_m, respectively, F_M is represented as $F_M = \{O_1, O_2, \dots, O_m\}|[t_i, t_j]$.

Note that d is a parameter distinguishing moving flock from flock, and Definition 5.5 is adopted as the definition of moving flock here and hereafter in this chapter.

5.2.2 Taxonomy of moving flock patterns

In this chapter, a brief hierarchical taxonomy of moving flock patterns is proposed in order to derive various types of moving flock patterns. According to the definitions of flock and moving flock, the main factors that can directly influence the classification of flock patterns are proximity (r), number of members in a flock (m), time duration (k) and displacement (d). Hence, the taxonomy is proposed based on these four factors. It is shown in Figure 5.3.

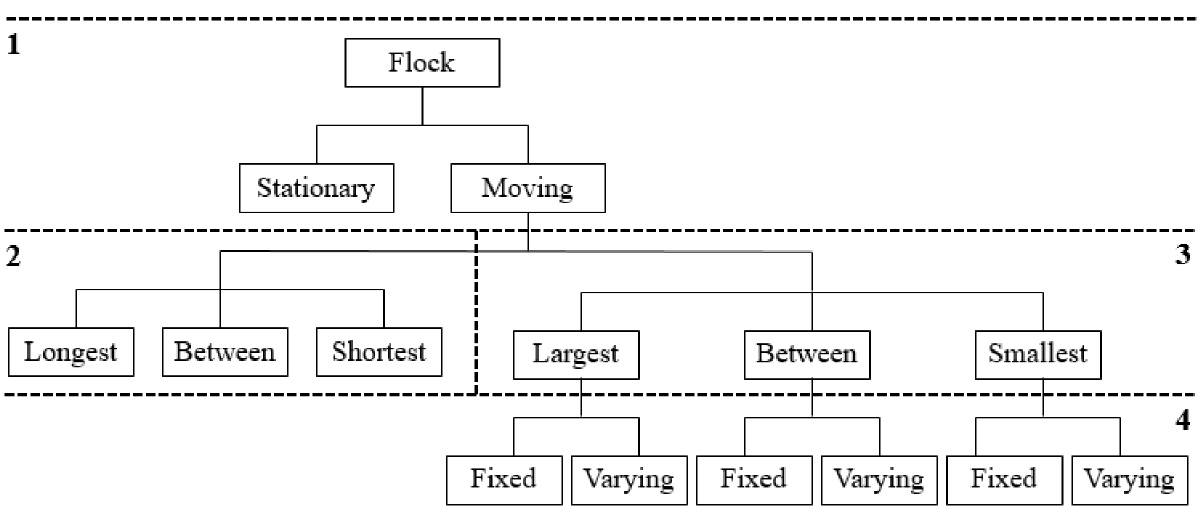


Figure 5.3. The taxonomy of moving flock patterns.

In Figure 5.3, the taxonomy can be classified into four main parts, i.e., parts 1, 2, 3 and 4. Part 1 is based on proximities and displacements, and it lies on the highest level. As a flock, the spatial locations of the involving members should be proximate at the same timestamp. When the displacement of locations meets the requirement of moving flock, it is classified as a moving flock pattern. Otherwise, stationary flock pattern. Parts 2 and 3 are based on time duration and number of members in a flock, respectively, and they are on the intermediate level. As for the time duration of a flock, it can be either the longest, the shortest or in between. Similarly, the number of members in a flock can be either the largest, the smallest or in between. Part 4 is on the lowest level, and it is based on the variability of flock members. The members in a moving flock can be either fixed or varying during the whole time interval.

According to the taxonomy, a large amount of types of moving flock patterns can be derived based on a sole part or the combinations of multiple parts. For example, three types of moving flock patterns can be derived based solely on part 3, which are longest duration moving flock patterns, shortest duration moving flock patterns and moving flock patterns with a time duration in between. Nevertheless, not all types of moving flock patterns are interesting to people in practice. Comparatively, the moving flock patterns that are with specific properties, such as the top p ($p \geq 1$) largest/smallest number of members, the top p ($p \geq 1$) longest/shortest time durations, and fixed flock members, might be more interesting to people. Note that as four parameters have already been involved in the definition of moving flock, the cases where p is larger than 1 are not considered in this chapter in order to reduce complexities. Hence, in this chapter, we are particularly interested in eight types (i.e., types A ~ H) of moving flock patterns which are derived based on one or mutiple of the following specific properties: the largest/smallest number of members, the longest/shortest time duration and fixed flock members. The eight types of moving flock patterns are listed in Table 5.1. Note that in Table 5.1, types A and B are derived based on part 2 in Figure 5.3, types C and D are derived based on part 3 in Figure 5.3, and types E, F, G and H are derived based on the combination (i.e., intersection) of one type from A or B and the other type from C or D. The detailed relationships between types E, F, G and H and types A, B, C and D are shown in Figure 5.4. The eight types of moving flock patterns are thus considered as the main focus in this chapter, and a Reeb graph-based approach is proposed in order to discover them.

Table 5.1. The eight types of moving flock patterns.

Type of moving flock patterns	Explanation
A	Moving flock patterns with longest time duration and fixed members
B	Moving flock patterns with shortest time duration and fixed members
C	Moving flock patterns with largest number of fixed members
D	Moving flock patterns with smallest number of fixed members
E	Moving flock patterns with longest time duration and largest number of fixed members
F	Moving flock patterns with longest time duration and smallest number of fixed members
G	Moving flock patterns with shortest time duration and largest number of fixed members
H	Moving flock patterns with shortest time duration and smallest number of fixed members

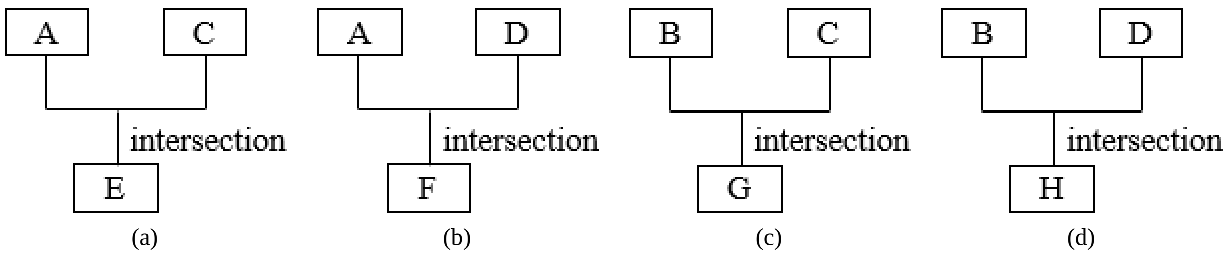


Figure 5.4. The relationships of moving flock patterns between: (a) type E and types A and B; (b) type F and types A and D; (c) type G and types B and C, and (d) type H and types B and D.

5.3 Methodology

In this chapter, we propose a Reeb graph-based approach to discover moving flock patterns. Specifically, we mainly aim to discover the eight types of moving flock patterns listed in Table 5.1. The approach includes five steps: (1) generating Reeb graphs based on movement data; (2) filtering specific Reeb graphs; (3) extracting flock patterns; (4) extracting moving flock patterns, and (5) extracting each type of moving flock patterns. In the following, the five steps are introduced in detail.

In order to make things easier to be understood, a sample dataset is illustrated to show how the approach works. The dataset is shown in Figure 5.5(a). Figure 5.5(a) illustrates the movements of five objects (note that the ids of the objects are 1, 2, 3, 4 and 5, respectively) moving during a time interval consisting of five consecutive timestamps (i.e., t_0, t_1, t_2, t_3 and t_4) and two other objects (the ids of the objects are 6 and 7) moving during a time interval consisting of two consecutive timestamps (i.e., t_2 and t_3). The discrete points denote the spatial locations of the objects at the corresponding timestamps.

5.3.1 Generating Reeb graphs based on movement data

Reeb graph is a concept from topology. It has already been extensively used in many fields, especially in shape analysis (Biasotti et al., 2008; Chen et al., 2013) and visualisation of scientific data (Fomenko & Kunii, 1997; Edelsbrunner & Harer, 2010). Given its strong ability in representing the evolution of connected components of the level sets on a two or higher dimensional scalar function, it has been introduced to trajectory analysis as well (Buchin et al. 2013; Zhang & Van de Weghe, 2018). Basically, it is a structure that reveals the temporal changes of the spatial closeness of a group of moving objects. Hence, it can be used to represent the evolution of grouping with time for moving objects. In this chapter, the Reeb graph is generated based on regularly sampled movement data. Similar to other types of graphs, the Reeb graph can also be represented as $= \{V, E\}$, where $V = \{v_1, v_2, \ldots, v_n\}$ and $E = \{v_i v_j\}$ $(1 \leq i \leq n, 1 \leq j \leq n)$ denote the set of vertices and edges, respectively. Generally, two steps are necessary for generating a Reeb graph: (1) the generation of vertices, and (2) the construction of edges. As on the one hand, the aim of generating the Reeb graphs is to discover moving flock patterns, and on the other hand, there are four key parameters (i.e., *r*, *m*, *k* and *d*) for a moving flock pattern, we propose to generate a simplified Reeb graph instead of a complete Reeb graph in order to reduce computational complexities. Specifically, a simplified Reeb graph is generated by deleting the vertices (and the corresponding edges that are connected to the vertices) from the corresponding complete Reeb graph. Note that for reasons of convenience, Reeb graph specifically denotes the simplified Reeb graph here and hereafter in this chapter. In general, three steps are included when generating simplified Reeb graphs: (1) the generation of vertices; (2) the deletion of specific vertices, and (3) the construction of edges.

5.3.1.1 The generation of vertices

As introduced previously, four parameters (i.e., *r*, *m*, *k* and *d*) are involved in a moving flock pattern. The generation of vertices is based on parameter *r*. For each timestamp, if the objects whose spatial locations at the same timestamp are within a circle of radius *r*, they are considered as being involved in the same vertex. Note that the selection of the base object (i.e., the object whose spatial location at a timestamp is considered as the centre of the corresponding circle) is important, as different base objects may result in different results. In this approach, we propose that for each timestamp, the object whose spatial distance (i.e., Euclidean distance, here and hereafter) is closest to the centroid of all the objects is considered as the base object. In this way, the base objects can be determined automatically, thus avoiding the variation of results caused by the random selection of base objects. The vertices of Reeb graphs are generated according to the four steps below:

(1) For each timestamp, calculate the centroid of the spatial locations of all objects;
(2) Find the object whose spatial distance is closest to the centroid, and consider this object as a base object;
(3) Find all the objects whose spatial distances to the base object are no more than *r*, and consider them together with the base object as being involved in the same vertex;
(4) For the remaining objects, recursively execute steps (1), (2) and (3) till every object at every timestamp is involved in a corresponding vertex. Thus, all vertices are generated.

Based on the proposed approach, for each timestamp, the objects can be divided into one or more groups. Note that a group means that the objects involved in the group are within a circle of radius *r*. Take Figure 5.5(a) for instance, assume *r* equals to a specific value, the groups of objects at each timestamp can be generated. The result is shown in Figure 5.5(b). The corresponding vertices of the Reeb graphs are shown in Figure 5.5(c), in which the points denote the vertices of the Reeb graphs and the numbers next to the vertices indicate the ids of the objects involved in the corresponding vertices.

5.3.1.2 The deletion of specific vertices

The deletion of specific vertices is executed based on parameter *m*. As there might be vertices in which the number of involved objects is less than *m*, we propose to delete such vertices. The method is described as follows:

(1) For each vertex, calculate the number of objects involved in it;

(2) If the number is less than *m*, then delete the vertex;

(3) Recursively execute steps (1) and (2) till all vertices are checked.

Take Figure 5.5(c) for example, assume *m* equals to 2, three vertices ought to be deleted in Figure 5.5(c). The result after deleting the three vertices is shown in Figure 5.5(d), in which all vertices involve at least two objects.

5.3.1.3 The construction of edges

When constructing the edges, two principles are obeyed: (1) any two vertices at the same timestamp are not allowed to be connected, and (2) any two vertices at two consecutive timestamps are connected if at least one common object is involved in both vertices. According to these principles, the edges of the vertices in Figure 5.5(d) can be constructed, which are displayed in Figure 5.5(e). Thus, the graphs in Figure 5.5(e) are considered as the final Reeb graphs generated based on the sample movement data.

5.3.2 Filtering specific Reeb graphs

The filtering of specific Reeb graphs is carried out based on parameter *k*, as the time durations of some Reeb graphs might be shorter than *k*. Hence, such Reeb graphs should be filtered. The method of filtering such Reeb graphs is described as follows:

(1) For each Reeb graph, find its minimum timestamp t_{min} and maximum timestamp t_{max}, if $t_{max} - t_{min} < k$, delete this graph;

(2) Recursively execute step (1) until all Reeb graphs are checked.

Take Figure 5.5(e) for example, assume *k* equals to 2, the Reeb graph on the right should be filtered. The result after filtering is shown in Figure 5.5(f).

5.3.3 Extracting flock patterns

The Reeb graphs existing so far all meet the requirements of flock, which means that for each Reeb graph: (1) the objects involved in each vertex are all within a circle of radius *r*; (2) the number of objects involved in each vertex is no less than *m*, and (3) the time duration of the Reeb graph is no less than k. Hence, based on each Reeb graph, one or multiple flock patterns can be extracted. The method for extracting flock patterns based on each Reeb graph is described in detail as follows:

(1) For a vertex v_i, find all subgroups (each of which consists of at least *m* objects) that can be formed by the objects involved in v_i;
(2) Select the first group, if this group hasn't been processed, then go to the next vertex v_{i+1} which connects to v_i. If this group can also be formed by the objects involved in v_{i+1}, then go to the next vertex which connects to v_{i+1}. Recursively execute this till this group cannot be formed by the objects involved in a vertex. Then, for this group, record the corresponding timestamps when it starts and ends. Besides, identify this group as processed;
(3) For the remaining groups formed by the objects involved in v_i, continue step (2) till all the groups are processed;
(4) For vertex v_{i+1}, repeat similar operations as demonstrated in steps (1), (2) and (3).

Based on the above operations, all groups of objects together with their corresponding time durations can be extracted. They can be considered as candidates of flock patterns. Nevertheless, among the extracted groups, some might no longer meet the requirement that the corresponding time durations should be no less than *k*. Hence, in order to extract the correct flock patterns, such groups have to be deleted. After deleting such groups, the remaining groups are considered as flock patterns.

Take Figure 5.5(f) for example, assume *m* still equals to 2 and *k* still equals to 2, five flock patterns can be extracted, which are $\{1,3\}|[t_0,t_3]$, $\{4,5\}|[t_1,t_4]$, $\{2,4\}|[t_2,t_4]$, $\{2,5\}|[t_2,t_4]$ and $\{2,4,5\}|[t_2,t_4]$, respectively.

5.3.4 Extracting moving flock patterns

Among the extracted flock patterns, however, not all of them are certainly moving flock patterns. Hence, each flock pattern has to be checked so that moving flock patterns can be correctly extracted. The method for extracting moving flock patterns is described below:

(1) For each flock pattern F_i, assume the corresponding time interval of F_i is $[t_0, t_n]$, calculate the spatial extent of F_i between t_j and t_{j+1} $(0 \leq j \leq n-1)$ $ext\left(F_i|t_j, F_i|t_{j+1}\right)$, if $ext\left(F_i|t_j, F_i|t_{j+1}\right) < d$ $(d > 0)$, then time interval $[t_0, t_n]$ is split into two sub time intervals, i.e., $[t_0, t_j]$ and $[t_{j+1}, t_n]$. Continue this until the spatial extents of F_i between any two consecutive timestamps are checked. Thus, F_i is divided into multiple sub flock patterns if there exits time intervals during which the objects involved in F_i appear stationary.

(2) Repetitively execute step (1) till all flock patterns are checked.

Based on the operations mentioned above, for specific flock patterns, a collection of sub flock patterns might be generated if there exists time intervals during which the involved objects are not 'moving'. However, among the generated sub flock patterns, the time durations of some might be shorter than *k*. In this case, such sub flock patterns have to be deleted. Finally, the remaining flock patterns are all considered as moving flock patterns.

Take Figure 5.5(b) for example, if we consider the spatial extent of flock pattern $\{1,3\}|[t_0, t_3]$ between t_2 and t_3 does not meet the requirement of a moving flock, then the flock pattern is split into two sub flock patterns: $\{1,3\}|[t_0, t_2]$ and $\{1,3\}|[t_3, t_3]$. Assume *m* still equals to 2 and *k* still equals to 2, $\{1,3\}|[t_3, t_3]$ is not considered as a flock pattern. Hence, based on the extracted flock patterns presented in section 3.3, we can finally extract all the moving flock patterns, which are $\{1,3\}|[t_0, t_2]$, $\{4,5\}|[t_1, t_4]$, $\{2,4\}|[t_2, t_4]$, $\{2,5\}|[t_2, t_4]$ and $\{2,4,5\}|[t_2, t_4]$, respectively.

5.3.5 Extracting the eight types of moving flock patterns

In this chapter, we mainly focus on the eight types of moving flock patterns listed in Table 5.1. The method to extract each of the eight types of moving flock patterns is described as follows:

(1) Assume *n* moving flock patterns $F_M = \{F_{M_1}, F_{M_2}, \ldots, F_{M_n}\}$, for each moving flock

pattern F_{M_i} ($1 \leq i \leq n$), calculate the number of objects involved in it and its corresponding time duration. Assume the results are n_i and t_i, respectively;

(2) Repetitively execute step (1) till all moving flock patterns are checked. The final results are stored in two collections N and T, respectively, where $N = \{n_1, n_2, \dots, n_n\}$ and $T = \{t_1, t_2, \dots, t_n\}$;

(3) Calculate the minimum values and the maximum values of *N* and *T*. Assume they are *min_N*, *max_N*, *min_T* and *max_T*, respectively;

(4) Each of the eight types of moving flock patterns is extracted according to the corresponding method as follows:

Type A: find out the moving flock patterns whose time duration equals to *max_T* in F_M;

Type B: find out the moving flock patterns whose time duration equals to *min_T* in F_M;

Type C: find out the moving flock patterns whose number of involved objects equals to *max_N* in F_M;

Type D: find out the moving flock patterns whose number of involved objects equals to *min_N* in F_M;

Type E: find out the moving flock patterns whose time duration equals to *max_T* and number of involved objects equals to *max_N* in F_M;

Type F: find out the moving flock patterns whose time duration equals to *max_T* and number of involved objects equals to *min_N* in F_M;

Type G: find out the moving flock patterns whose time duration equals to *min_T* and number of involved objects equals to *max_N* in F_M;

Type H: find out the moving flock patterns whose time duration equals to *min_T* and number of involved objects equals to *min_N* in F_M.

Based on this method, all the eight types of moving flock patterns can be extracted. Take the sample dataset in Figure 5.5 for instance, the finally extracted eight types of moving flock patterns are listed in Table 5.2.

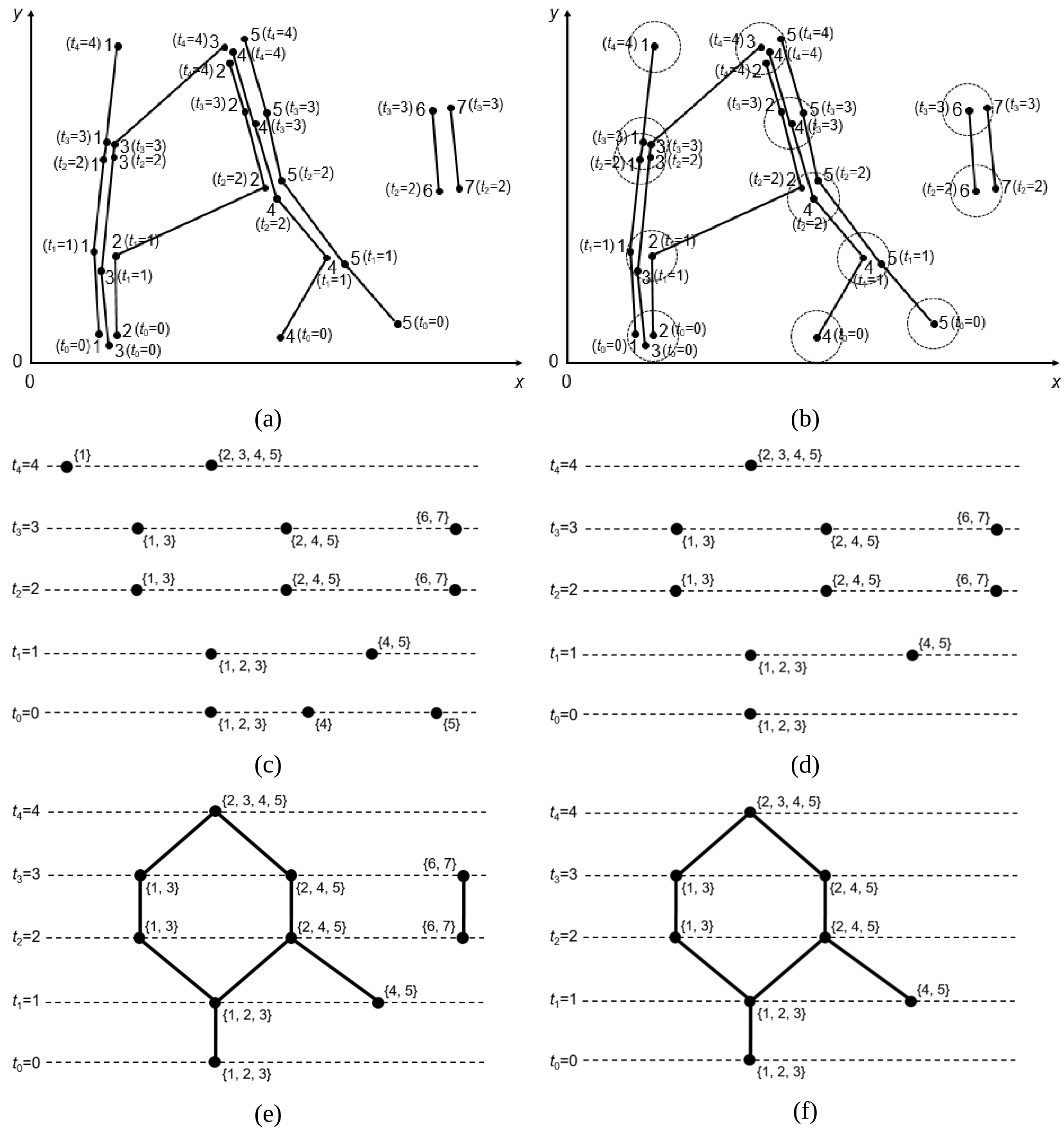


Figure 5.5. Illustration of the methodology: (a) the sample dataset, (b) the groups of objects at each timestamp according to parameter *r*, (c) the generated vertices, (d) the remaining vertices after deleting specific ones, (e) the generated Reeb graphs, and (f) the remaining Reeb graph after filtering the specific one.

Table 5.2. The extracted eight types of moving flock patterns based on the sample dataset shown in Figure 5.4.

Type of moving flock patterns	Extracted moving flock patterns
A	$\{4, 5\}\|[t_1, t_4]$
B	$\{1, 3\}\|[t_0, t_2]$
	$\{2, 4\}\|[t_2, t_4]$
	$\{2, 5\}\|[t_2, t_4]$
	$\{2, 4, 5\}\|[t_2, t_4]$
C	$\{2, 4, 5\}\|[t_2, t_4]$
D	$\{1, 3\}\|[t_0, t_2]$
	$\{4, 5\}\|[t_1, t_4]$
	$\{2, 4\}\|[t_2, t_4]$
	$\{2, 5\}\|[t_2, t_4]$
E	none
F	$\{4, 5\}\|[t_1, t_4]$
G	$\{2, 4, 5\}\|[t_2, t_4]$
H	$\{1, 3\}\|[t_0, t_2]$
	$\{2, 4\}\|[t_2, t_4]$
	$\{2, 5\}\|[t_2, t_4]$

5.4 Case study

5.4.1 Dataset

The movement data adopted in this chapter were obtained from a real and entire football match between 'Club Brugge KV' and 'Standard de Liège' which took place on 2nd March 2014. For simplicity, we call them 'Club Brugge' and 'Standard Liège' respectively in the remainder of this chapter. The dataset includes both spatio-temporal information and semantic information. The spatio-temporal information is recorded in a (id, x, y, t) format, where *id* identifies a specific player, x and y respectively denote the x and y coordinates of the player's position, and t represents the corresponding timestamp. The semantic information mainly includes the basic information of both teams (such as names of players, *id* numbers of players and positions played) and the events that happened during the match (such as event name, time of occurrence and *id*s of the actors). The spatio-temporal information can be used to discover moving flock patterns based on the approach proposed in this chapter, and the semantic information can be used to validate the discovered moving flock patterns when necessary.

As football is considered as a highly interactive sport (since the players need to interact frequently with the teammates), various types of interaction patterns can be involved. Among them, moving flock patterns are the ones that are of particular interest to us. This is important as the discovery and exploration of moving flock patterns can give insight into a team's playing style, thus might be useful for potential tactics arrangement. In this chapter, as the main focus is on the methodology, we only use the movement data of the players (except the goalkeeper) of Club Brugge during the first five minutes (i.e., [0, 300] s). This is because coaches might be very interested in the opening of the match, as they might give their instructions on the tactics just a few minutes before the start of the match, and want to evaluate them in the first minutes of the match. In the original dataset, the locations of all the players were tracked at a temporal resolution of 0.1 s. In order to reduce computational complexities, we down-sampled the temporal resolution from 0.1 s to 1 s. After processing, 3010 discrete points and 10 trajectories are generated. They are visualised in Figure 5.6. In addition, the ten players are represented by player 1, player 2, player 3, …, and player 10 due to privacy issues.

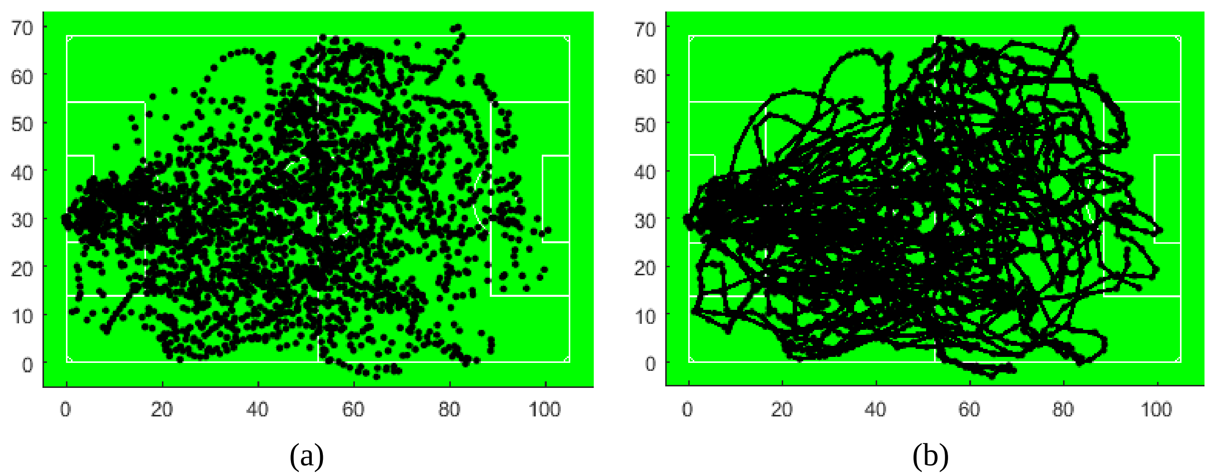


Figure 5.6. Visualisation of the movement data: (a) the discrete points, and (b) the trajectories.

5.4.2 Results and analysis

As is introduced, four parameters (i.e., *r*, *m*, *k* and *d*) are involved in the proposed approach, hence, different parameter values might result in the variation of moving flock patterns discovered. Based on this, the relations between the values of parameters and the number of moving flock patterns discovered are explored. Note that when exploring the relations between the varying values of one sole parameter and the number of moving flock patterns discovered, the other three parameters are set to the default values. The default values of the four parameters are listed in Table 5.3. The reason why considering them as default values is that: (1) for each parameter, the default value is neither too large nor too small, and (2) relatively more moving flock patterns can be found, based on which much useful information might be provided. Note that the units of *r*, *k* and *d* are meter, second and meter, respectively. The relations between the values of the four parameters and the number of moving flock patterns discovered are shown in Figure 5.7.

Table 5.3. The default values of the four parameters.

The parameters	The default values
r	10
m	3
k	3
d	0.5

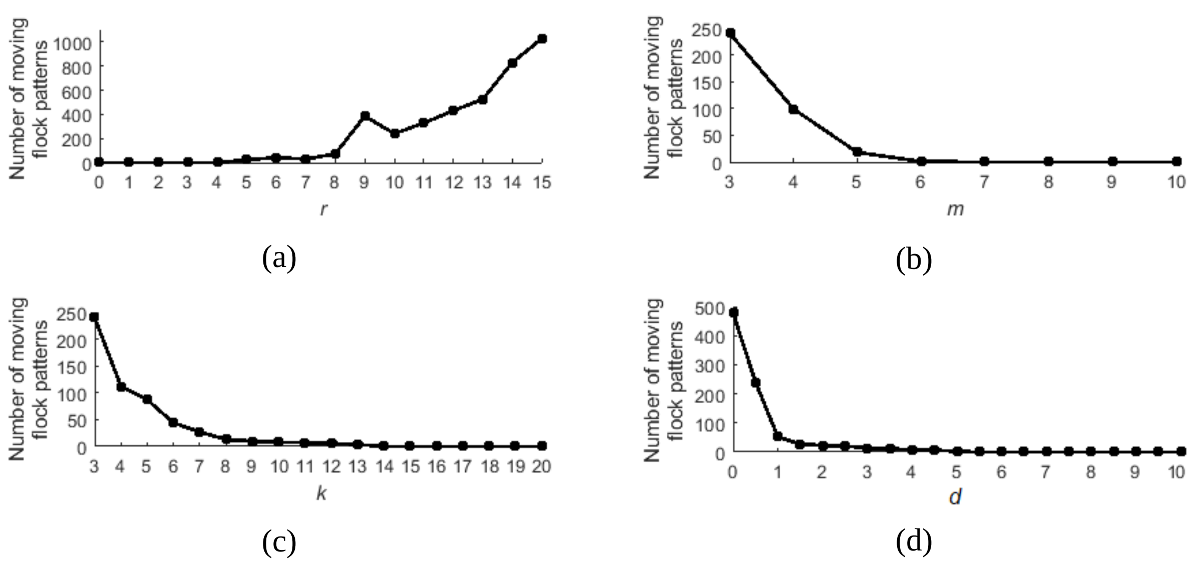


Figure 5.7. The relations between the values of parameters and the number of moving flock patterns discovered: (a) *r*, (b) *m*, (c) *k*, and (d) *d*.

According to Figure 5.7, we can find that for parameter *r*, the number of moving flock patterns becomes larger when *r* becomes larger in general. However, there is an outlier when *r* equals to 9. Hence, 9 appears to be a good value for *r*. For parameters *m*, *k* and *d*, the number of moving flock patterns becomes smaller when the values become larger.

However, when the values are larger than specific values, no any moving flock pattern can be discovered. According to this, we can find that the default values for *m*, *k* and *d* listed in Table 5.3 appear reasonable. In addition, the figures reveal that the impacts of *m*, *k* and *d* on the results appear more regular and predictable compared to that of *r*.

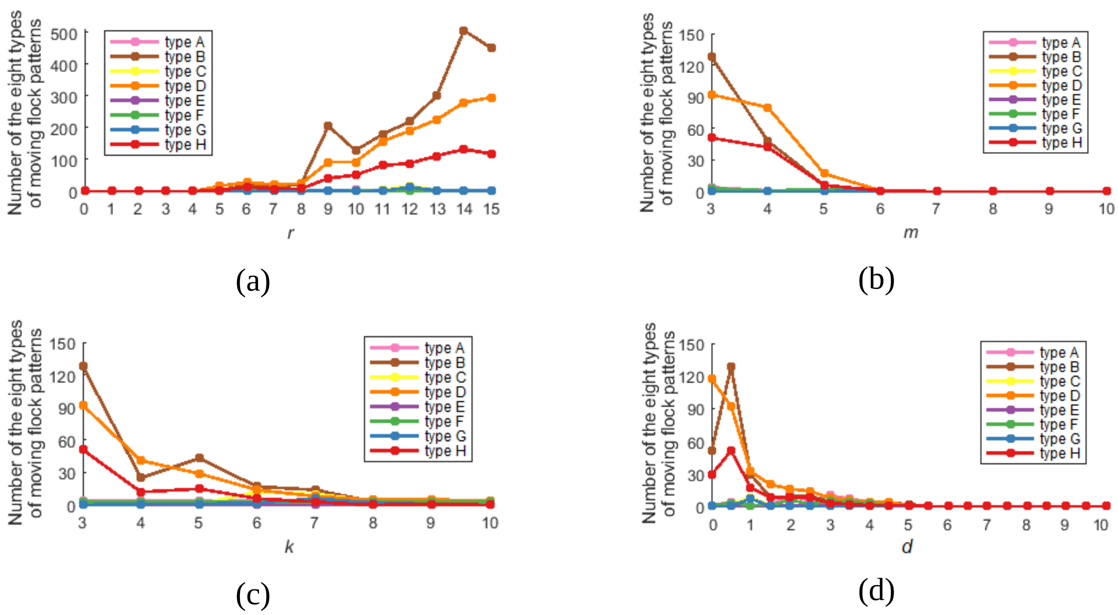


Figure 5.8. The relations between the values of parameters and the number of the eight types of moving flock patterns discovered: (a) *r*, (b) *m*, (c) *k*, and (d) *d*.

In addition, the relations between the values of parameters and the number of the eight types of moving flock patterns discovered (i.e., types A ~ H) are explored as well, which are shown in Figure 5.8. From Figure 5.8, we can notice that generally the number of five types of moving flock patterns are relatively large, which are types B, D, F, G and H. Comparatively, the number of the other types of moving flock patterns is quite small. Besides, the overall trends between the number of moving flock patterns of types D, G and H and parameters *r*, *m* and *d* are similar to that in Figure 5.7. However, slight differences still exist. For moving flock patterns of type B, the differences are even obvious. This demonstrates that the number of the eight types of moving flock patterns discovered does not obey strict rules with the number of moving flock patterns discovered, although the overall trends are similar. On the other hand, for types A, C, E, and F, there are little differences between the number of moving flock patterns discovered and the values of the four parameters. This is because types E, F, G and H can be considered as subsets of either type A, type B, type C or type D (as shown in Figure 5.4). Hence, the number of moving flock patterns which are considered as subsets is comparatively smaller.

In addition to the values of individual parameters, the combination of different parameter vales can also influence the number of moving flock patterns discovered. Hence, key to this is to find a suitable value for each parameter. In order to achieve this, we perform the hierarchical clustering of the four parameters based on Figure 5.7. This is because more details (e.g., similarities and dissimilarities) can be revealed level by level via hierarchical clustering, based on which suggestions on good parameter values can be provided. Thus, one can determine the optimal parameter values according to his/her specific demands. The result after hierarchical clustering is shown in Figure 5.9. In Figure 5.9, for each dendrogram, the height of the links between/among the elements in the same cluster denotes their degree of similarity. A higher link demonstrates that the corresponding elements are more dissimilar to each other. The optimal parameter values thus can be determined based on the degree of similarity one desires. In our case, for each parameter, we determine the value who has the most dissimilarities with others to be the optimal value. Based on this, we can conclude that the optimal values for parameters m, k, and d are 4, 3, and 0.5, respectively. For parameter r, although values of 14 and 15 have the largest dissimilarity, they are not considered as the optimal values. This is because, according to the relations demonstrated in Figure 5.7(a), the number of moving flock patterns is rather too large when r equals to 14 or 15, thus the result appears not so satisfying. Hence, the values at a lower dissimilarity level (i.e., 9 and 12) appears fine. Combined with the outlier revealed in Figure 5.7(a), we adopt 9 as the optimal value for r. Therefore, according to Figure 5.9, the optimal values for the four parameters are: $r = 9$, $m = 4$, $k = 3$ and $d = 0.5$. We hence try the proposed approach under the following parameter combination: $r = 9$, $m = 4$, $k = 3$ and $d = 0.5$. Finally, 72 moving flock patterns are discovered. As not all of them are interesting, we further discover the eight types of moving flock patterns. In particular, the number of the eight types of moving flock patterns discovered are shown in Table 5.4.

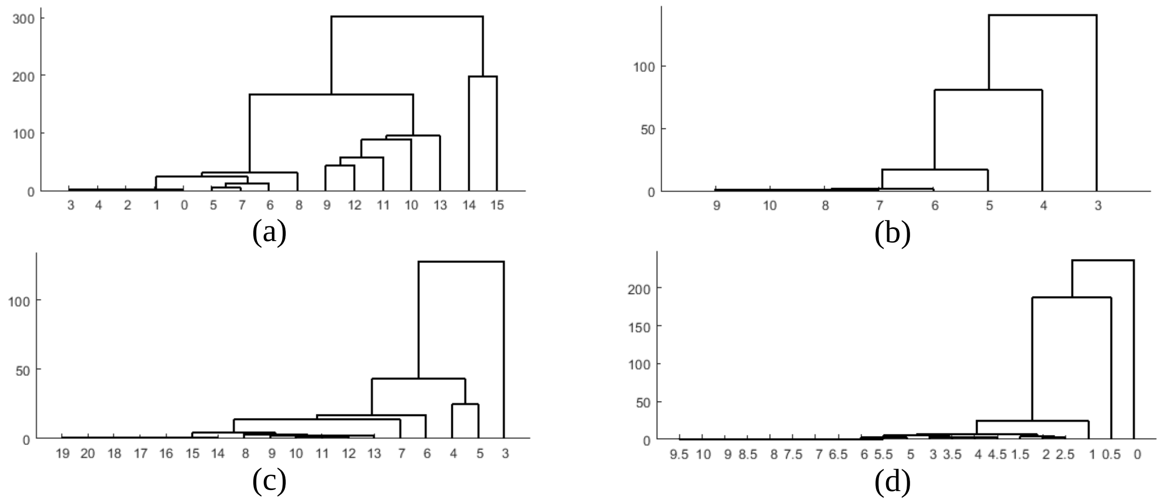


Figure 5.9. The hierarchical clustering of the four parameters: (a) *r*; (b) *m*; (c) *k*, and (d) *d*.

Table 5.4. The number of the eight types of moving flock patterns discovered based on the proposed approach under the parameter combination: $r = 9$, $m = 4$, $k = 3$ and $d = 0.5$.

Type of moving flock patterns	Number of moving flock patterns
A	4
B	33
C	1
D	40
E	0
F	3
G	1
H	17

From Table 5.4, we can see that the numbers of some specific types of moving flock patterns are quite large, such as types D and B. This demonstrates that among the moving flock patterns, a large part is with the shortest time duration (type B) or with the smallest number of objects (type D). Comparatively, the number of moving flock patterns which are with the longest duration (type A) or with the largest number of objects (type C) are quite small. Furthermore, there is even no such pattern which is with both the longest duration and the largest number of objects (type E). In order to investigate the discovered moving flock patterns deeply, we exhibit the detailed information (i.e., the involved players and the corresponding time interval) of each moving flock pattern. Note that for illustration purposes and due to the length of the chapter, we only show the first three moving flock patterns of types A, B, D and H. The detailed information is listed in Table 5.5. The moving flock patterns which are listed in Table 5 are visualised in Figure 5.10.

Table 5.5. The detailed information of the eight types of illustrated moving flock patterns under the parameter combination: r = 9, m = 4, k = 3 and d = 0.5.

Type of moving flock patterns	Detailed information of the moving flock patterns
A	{2, 8, 9, 10}|[244, 259]
B	{1, 8, 9, 10}|[264, 268]
	{2, 3, 9, 10}|[264, 268]
	{2, 5, 7, 10}|[247, 251]
C	{1, 4, 5, 8, 9, 10}|[247, 252]
D	{3, 5, 7, 9}|[233, 238]
	{1, 5, 8, 9}|[247, 252]
	{1, 8, 9, 10}|[244, 252]
E	none
F	{2, 8, 9, 10}|[244, 259]
G	none
H	{1, 8, 9, 10}|[264, 268]
	{2, 3, 9, 10}|[264, 268]
	{2, 5, 7, 10}|[247, 251]

From Table 5.5, we can observe that generally there are only seven different moving flock patterns, as some of them affiliate to multiple types. Take moving flock pattern {5, 8, 9, 10}|[244, 252] for example, it can be considered as either type A or type F. According to Figure 5.10 we can see that generally the overall variations of the spatial locations of the eight moving flock patterns listed in Table 5.5 are not quite large. By referring to the semantic information of the dataset, we find that the match was temporarily interrupted (e.g., 'clearance' event happened) during these time intervals, which hence resulted in the relatively small variations of spatial locations, as the players were not running fast. One important reason for this might be that 0.5 is not a quite large value for parameter d. Hence, some of the moving flock patterns which the spatial extents between two consecutive timestamps are not quite large are involved. In all, this demonstrates that the proposed approach indeed has the capability to discover moving flock patterns and further distinguish the eight types of moving flock patterns.

In order to validate the proposed approach further, we also try another parameter combination: r = 15, m = 4, k = 3 and d = 2. This is because under this parameter combination, on the one hand, more number of moving flock patterns might be discovered. On the other hand, the variations of spatial locations of the discovered moving flock

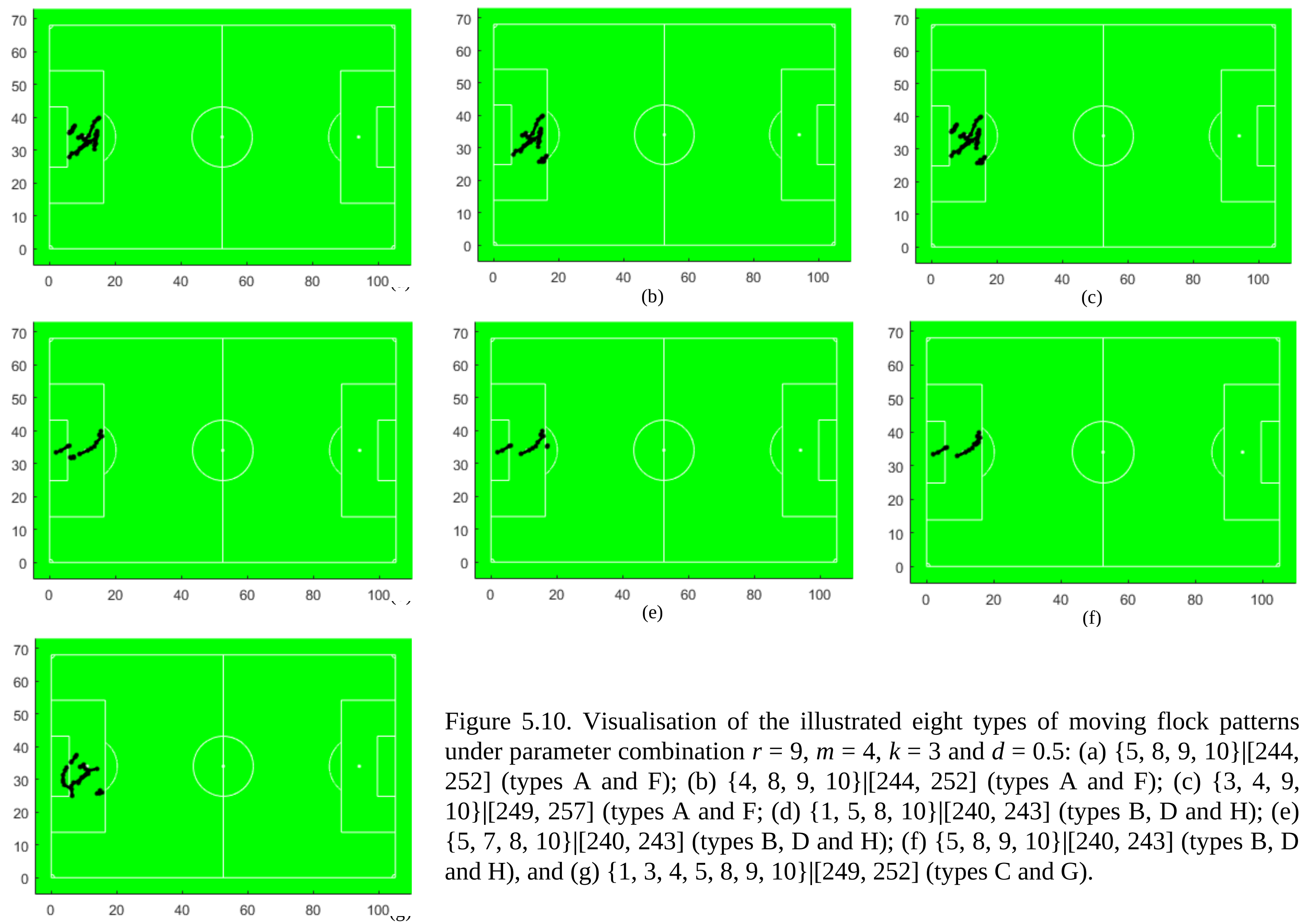


Figure 5.10. Visualisation of the illustrated eight types of moving flock patterns under parameter combination $r = 9$, $m = 4$, $k = 3$ and $d = 0.5$: (a) {5, 8, 9, 10}|[244, 252] (types A and F); (b) {4, 8, 9, 10}|[244, 252] (types A and F); (c) {3, 4, 9, 10}|[249, 257] (types A and F; (d) {1, 5, 8, 10}|[240, 243] (types B, D and H); (e) {5, 7, 8, 10}|[240, 243] (types B, D and H); (f) {5, 8, 9, 10}|[240, 243] (types B, D and H), and (g) {1, 3, 4, 5, 8, 9, 10}|[249, 252] (types C and G).

patterns should be larger, which might be more interesting to people. The results show that 94 moving flock patterns are discovered under this parameter combination. Specifically, the number of the eight types of moving flock patterns and the detailed information of the illustrated moving flock patterns under this parameter combination are listed in Table 5.6 and Table 5.7, respectively. The moving flock patterns listed in Table 5.7 are visualised in Figure 5.11.

Table 5.6. The number of the eight types of moving flock patterns discovered based on the proposed approach under the parameter combination: r = 15, m = 4, k = 3 and d = 2.

Type of moving flock patterns	Number of moving flock patterns
A	1
B	34
C	6
D	60
E	0
F	1
G	4
H	18

Table 5.7. The detailed information of the eight types of illustrated moving flock patterns under the parameter combination: r = 15, m = 4, k = 3 and d = 2.

Type of moving flock patterns	Detailed information of the moving flock patterns
A	{1, 2, 6, 7}|[175, 182]
B	{1, 4, 5, 7}|[14, 17]
	{1, 4, 7, 8}|[15, 18]
	{2, 4, 7, 8}|[15, 18]
C	{1, 2, 4, 7, 8, 9}|[15, 18]
	{1, 2, 4, 7, 8, 10}|[15, 18]
	{1, 2, 4, 7, 9, 10}|[15, 21]
D	{1, 5, 7, 10}|[6, 11]
	{1, 2, 4, 7}|[15, 21]
	{1, 2, 4, 9}|[15, 21]
E	none
F	{1, 2, 6, 7}|[175, 182]
G	{1, 2, 4, 7, 8, 9}|[15, 18]
	{1, 2, 4, 7, 8, 10}|[15, 18]
	{1, 4, 7, 8, 9, 10}|[15, 18]
H	{1, 4, 5, 7}|[14, 17]
	{1, 4, 7, 8}|[15, 18]
	{2, 4, 7, 8}|[15, 18]

From Table 5.7, we can see that generally eleven different moving flock patterns are listed. From Figure 5.11, we can find that the overall variations of the spatial locations of the eleven moving flock patterns are larger than that visualised in Figure 5.10. By combining

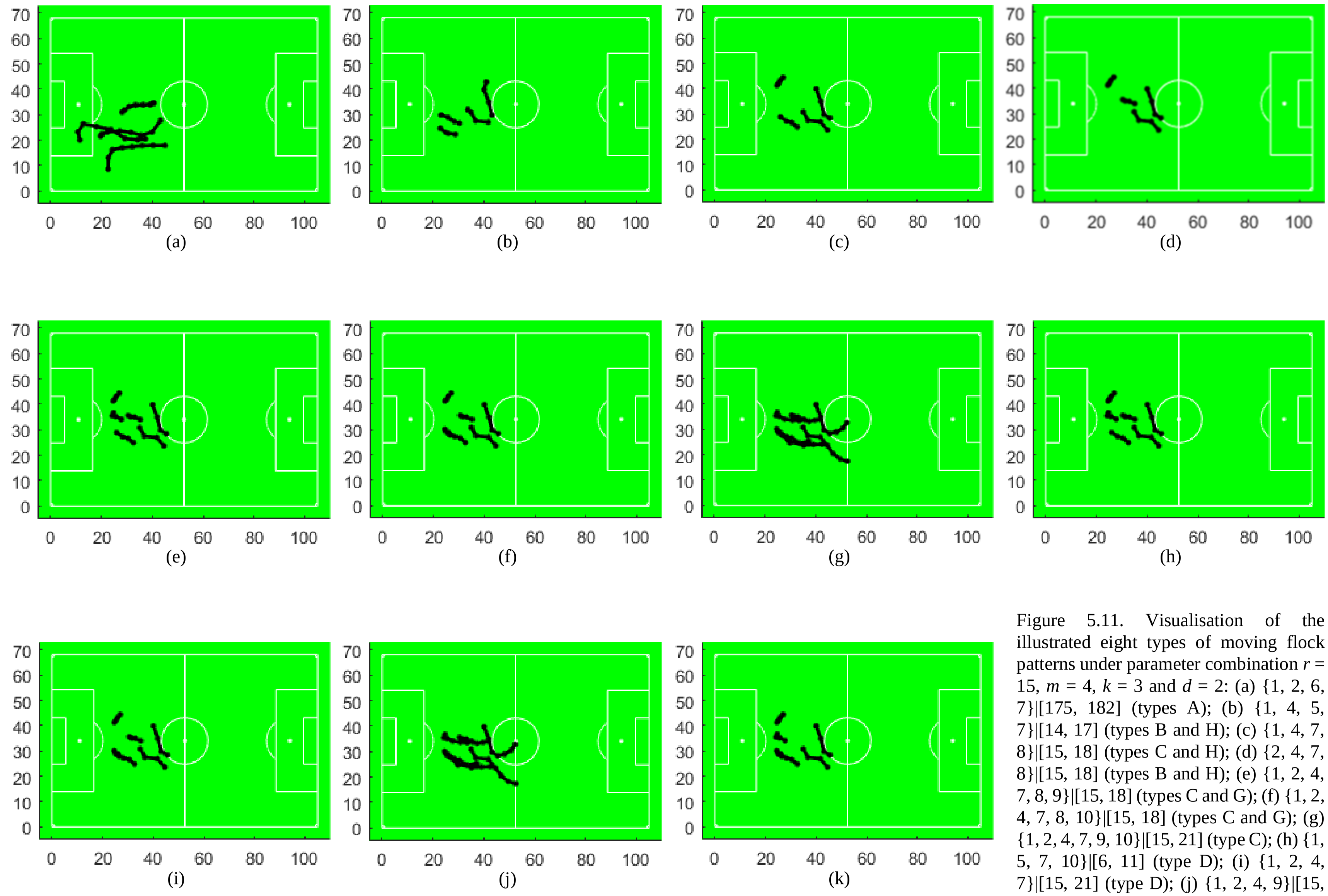


Figure 5.11. Visualisation of the illustrated eight types of moving flock patterns under parameter combination r = 15, m = 4, k = 3 and d = 2: (a) {1, 2, 6, 7}|[175, 182] (types A); (b) {1, 4, 5, 7}|[14, 17] (types B and H); (c) {1, 4, 7, 8}|[15, 18] (types C and H); (d) {2, 4, 7, 8}|[15, 18] (types B and H); (e) {1, 2, 4, 7, 8, 9}|[15, 18] (types C and G); (f) {1, 2, 4, 7, 8, 10}|[15, 18] (types C and G); (g) {1, 2, 4, 7, 9, 10}|[15, 21] (type C); (h) {1, 5, 7, 10}|[6, 11] (type D); (i) {1, 2, 4, 7}|[15, 21] (type D); (j) {1, 2, 4, 9}|[15, 21] (type D), and (k) {1, 4, 7, 8, 9, 10}|[15, 18] (type G).

the semantic information, we found that during the time intervals listed in Table 5.7, the player were either passing the ball, receiving the ball or running with the ball. Hence, the spatial locations of the moving flock patterns vary more. According to Figure 5.10 and Figure 5.11, we can conclude that the proposed approach can indeed be used to discover moving flock patterns, and different results can be derived under different parameter combinations. One can thus select specific parameter values according to his/her own demands when necessary in order to find the desired moving flock patterns.

Table 5.8. The top five most frequently appeared groups of players in all the detected moving flock patterns under the parameter combination: r = 9, m = 4, k = 3 and d = 0.5.

Groups of players involved in moving flock patterns	Number of appearance
{3, 4, 8, 10}	13
{3, 4, 9, 10}	12
{1, 3, 4, 10}	12
{1, 3, 4, 9}	12
{3, 4, 8, 9}	11

Table 5.9. The top five most frequently appeared groups of players in all the detected moving flock patterns under the parameter combination: r = 15, m = 4, k = 3 and d = 2.

Groups of players involved in moving flock patterns	Number of appearance
{4, 7, 9, 10}	10
{4, 7, 8, 10}	10
{4, 7, 8, 9}	10
{2, 4, 7, 10}	10
{2, 4, 7, 9}	10

Moreover, the group of players which are involved in the same moving flock pattern for multiple times appears interesting, as this group might interact well with each other comparatively. Thus, it might be helpful for coaches to arrange specific tactics based on this. We also provide an additional function to detect such groups of players from the discovered moving flock patterns. For illustration purposes, the top five groups of players which appear frequently in the discovered moving flock patterns under the two aforementioned parameter combinations are listed in Table 5.8 and Table 5.9, respectively. In Table 5.8 and Table 5.9, the first column denotes the group of players and the second column corresponds to the number of appearances of the group. Suggestions might be provided to sports professionals (e.g., coaches) for potential tactical arrangements based on the results.

5.5 Conclusions and future work

Currently, movement data are collected in a variety of domains, thus are becoming a popular type of data. Many research topics have been undertaken with respect to movement data, among which the discovery of various movement patterns takes a large proportion. As key contributions of this chapter, first, an improved definition of moving flock is developed based on existing definitions. Second, a taxonomy of moving flock patterns is proposed, based on which eight types of interesting moving flock patterns are derived. Third, a Reeb graph based approach is proposed to discover moving flock patterns, and the approach is further used to discover the eight types of moving flock patterns. In general, five steps are included in the proposed approach: (1) the corresponding Reeb graphs are generated based on the movement data; (2) the Reeb graphs which do not meet specific requirements are filtered; (3) flock patterns are extracted from the remaining Reeb graphs; (4) moving flock patterns are discovered based on the extracted flock patterns, and (5) each of the eight types of moving flock patterns is discovered based on the extracted moving flock patterns. The approach is then validated based on the movement data obtained from a real football match. The results demonstrate that the proposed approach can indeed be used to discover moving flock patterns, and potential insights can be provided to sports professionals for tactics arrangements.

In this chapter, when generating the Reeb graphs, only the simplified Reeb graphs are considered. The complete Reeb graphs can also be considered in the future. This is because with the complete Reeb graphs, the entire changes (e.g., spatial configuration) of all moving objects over time can be investigated, according to which more insightful information can be provided. With respect to the proposed approach, it involves four parameters. The impacts of the four parameters on the results are analysed. Although we suggest a way to select suitable parameter values, the selection of optimal parameter values is still an open question. In spite that the proposed approach is validated using the football movement data obtained from a real football match and the approach shows its effectiveness, it is still a relatively small dataset. Larger datasets (e.g., the movement data of the whole match) might be adopted in the future to discover more moving flock patterns in order to provide more insights to sports professionals when necessary. In addition, the movement data from other domains can be employed to extend the range of applications of the proposed approach.

References

Andersson, M., Gudmundsson, J., Laube, P., & Wolle, T. (2008). Reporting leaders and followers among trajectories of moving point objects. *GeoInformatica*, *12*(4), 497-528.

Benkert, M., Gudmundsson, J., Hübner, F., & Wolle, T. (2008). Reporting flock patterns. *Computational Geometry*, *41*(3), 111-125.

Biasotti, S., Giorgi, D., Spagnuolo, M., & Falcidieno, B. (2008). Reeb graphs for shape analysis and applications. *Theoretical Computer Science*, *392*, 5-22.

Buchin, K., Buchin, M., van Kreveld, M., Speckmann, B., & Staals, F. (2013). Trajectory grouping structure. In *Proceedings of the Workshop on Algorithms and Data Structures* (pp. 219-230).

Cao, Y., Zhu, J., & Gao, F. (2016). An algorithm for mining moving flock patterns from pedestrian trajectories. *Lecture Notes in Computer Science*, *9865*, 310-321.

Chen, F., Obermaier, H., Hagen, H., Hamann, B., Tierny, J., & Pascucci, V. (2013). Topology analysis of time-dependent multi-fluid data using the Reeb graph. *Computer Aided Geometric Design*, *30*(6), 557-566.

Civilis, A., Jensen, C. S., & Pakalnis, S. (2005). Techniques for efficient road-networkbased tracking of moving objects. *IEEE Transactions on Knowledge and Data Engineering*, *17*(5), 698-712.

Delafontaine, M., Versichele, M., Neutens, T., & Van de Weghe, N. (2012). Analysing spatiotemporal sequences in bluetooth tracking data. *Applied Geography*, *34*, 659-668.

Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Van de Weghe, N., et al. (2015). Analysis and visualisation of movement: An interdisciplinary review.

Movement Ecology, *3*(1), 5.

Dodge, S., Weibel, R., & Lautenschütz, A. K. (2008). Towards a taxonomy of movement patterns. *Information visualization*, *7*(3-4), 240-252.

Edelsbrunner, H., & Harer, J. L. (2010). *Computational topology: an introduction*. American Mathematical Society.

Eppstein, D., Goodrich, M., & Sun, J. (2005). The skip quadtree: A simple dynamic data structure for multidimensional data. In *Proceedings of the 21st ACM symposium on computational geometry* (pp. 296-305).

Fomenko, A., & Kunii, T. (1997). *Topological methods for visualization*. Springer, Tokyo, Japan.

Fort, M., Sellarès, J. A., & Valladares, N. (2014). A parallel GPU-based approach for reporting flock patterns. *International Journal of Geographical Information Science*, *28*(9), 1877-1903.

Gomez, G., López, P.H., Link, D., & Eskofier, B. (2014). Tracking of ball and players in beach volleyball videos. *PloS one*, *9*, e111730.

Gudmundsson, J., & van Kreveld, M. (2006). Computing longest duration flocks in trajectory data. In *Proceedings of the 14th annual ACM international symposium on advances in geographic information systems* (pp. 35-42).

Gudmundsson, J., van Kreveld, M., & Speckmann, B. (2007). Efficient detection of patterns in 2D trajectories of moving points. *Geoinformatica*, *11*(2), 195-215.

Gudmundsson, J., & Wolle, T. (2014). Football analysis using spatio-temporal tools. *Computers, Environments and Urban Systems*. *47*, 16-27.

Jacob, G. M., & Idicula, S. M. (2012). Detection of flock movement in spatio-temporal database using clustering techniques-An experience. In *Proceedings of the 2012 IEEE International Conference on Data Science & Engineering* (pp. 69-74).

Jeung, H., Yiu, M. L., Zhou, X., Jensen, C. S., & Shen, H. T. (2008). Discovery of convoys in trajectory databases. In *Proceedings of the VLDB Endowment*, (pp. 1068-1080).

Kjærgaard, M. B., Wirz, M., Roggen, D., & Tröster, G. (2012). Mobile sensing of pedestrian flocks in indoor environments using wifi signals. In *Proceedings of the*

2012 IEEE International Conference on Pervasive Computing and Communications (pp. 95-102).

Kalnis, P., Mamoulis, N., & Bakiras, S. (2005). On discovering moving clusters in spatiotemporal data. In *International Symposium on Spatial and Temporal Databases* (pp. 364-381).

Laube, P., & Imfeld, S. (2002). Analyzing relative motion within groups of trackable moving point objects. *Lecture notes in computer science*, *2478*, 132-144.

Laube, P., Imfeld, S., & Weibel, R. (2005). Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science*, *19*(6), 639-668.

Laube, P., van Kreveld, M., & Imfeld, S. (2004). Finding REMO – detecting relative motion patterns in geospatial lifelines. In *Proceedings of the 11th international symposium on spatial data handling* (pp. 201-214).

Lee, J., Han, J., & Whang, K. (2007). Trajectory clustering: A partition-and-group framework. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data* (pp. 593-604).

Loglisci, C. (2017). Using interactions and dynamics for mining groups of moving objects from trajectory data. *International Journal of Geographical Information Science*, 1-33.Shamoun-Baranes, J., van Loon, E. E., Purves, R. S., Speckmann, B., Weiskopf, D., & Camphuysen, C. J. (2012). Analysis and visualization of animal movement. *Biology Letters*, *8*(1), 6-9.

Shamoun-Baranes, J., van Loon, E. E., Purves, R. S., Speckmann, B., Weiskopf, D., & Camphuysen, C. J. (2012). Analysis and visualization of animal movement. *Biology Letters*, *8*(1), 6-9.

Solera, F., Calderara, S., & Cucchiara, R. (2015). Learning to identify leaders in crowd. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 43-48).

Turdukulov, U., Calderon Romero, A. O., Huisman, O., & Retsios, V. (2014). Visual mining of moving flock patterns in large spatio-temporal data sets using a frequent pattern approach. *International Journal of Geographical Information Science*, *28*(10), 2013-2029.

Vieira, M. R., Bakalov, P., & Tsotras, V. J. (2009). On-line discovery of flock patterns in spatio-temporal data. In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems* (pp. 286-295).

Wachowicz, M., Ong, R., Renso, C., & Nanni, M. (2011). Finding moving flock patterns among pedestrians through collective coherence. *International Journal of Geographical Information Science*, *25*(11), 1849-1864.

Wood, Z., & Galton, A. (2009). A taxonomy of collective phenomena. *Applied Ontology*, *4*, 267-292.

Yeoman, J., & Duckham, M. (2016). Decentralized detection and monitoring of convoy patterns. *International Journal of Geographical Information Science*, *30*(5), 993-1011.

Zhang, L., & Van de Weghe, N. (2018). Attribute trajectory analysis: a framework to analyse attribute changes using trajectory analysis techniques. *International Journal of Geographical Information Science*, 1-17.

Zhang, P., Beernaerts, J., & Van de Weghe, N. (2018). A hybrid approach combining the multi-temporal scale spatio-temporal network with the continuous triangular model for exploring dynamic interactions in movement data: a case study of football. *ISPRS International Journal of Geo-Information*, *7*(1), 31.

Zhang, P., Beernaerts, J., Zhang, L., & Van de Weghe, N. (2016). Visual exploration of match performance based on football movement data using the Continuous Triangular Model. *Applied Geography*, *76*, 1-13.

Zhang, P., Deng, M., Shi, Y., & Zhao, L. (2017). Detecting hotspots of urban residents' behaviours based on spatio-temporal clustering techniques. *GeoJournal*, *82*(5), 923-935.

6

General Discussion and Conclusions

6.1 General discussion

This section discusses the main outcomes and contributions of this thesis within a general research background. Because each of the chapters in this thesis focuses on at least one of the four research questions proposed in Chapter 1, this section elaborates on the value of this research and the remaining issues in a number of respects. Recommendations for future work are proposed where appropriate.

As introduced in previous chapters, the amount of movement data is increasing with the rapid development of location-aware techniques. Thus, these data are among the most frequent and important data sources today, especially in the big data era. Movement data usually include abundant information, such as spatio-temporal information and semantic information, in which important information/knowledge with respect to one or multiple moving object(s) is hidden. Because moving objects can denote nearly all types of objects in the world ranging from the objects with huge volumes (e.g., planets) to the objects with extremely small sizes (e.g., molecules), the analysis of movement data has broad applications in various domains. The aim of this thesis is to develop new approaches for analysing movement data from the perspective of geographical information science (GIScience). We aim to propel the development of research with respect to the analytic approaches or techniques of movement data in the GIScience domain so that voluminous and various movement data can be easily analysed instantaneously or gradually in the future. In addition, one important innovation of this thesis is the application of state-of-the-art approaches of analysing movement data into one novel field (i.e., sports), which has received increasing attention recently, to provide added value to current research in sports analytics.

We first aim to conduct a visual analysis of movement data. Compared with previous research, we proposed the application of the Continuous Triangular Model (CTM) to the visual analysis of movement data. The CTM is a novel model that was originally developed by Qiang et al. (2014). The most distinctive characteristic of the CTM is that it can represent continuous changes of information for objects at all temporal scales and with good visualisation capability. Accordingly, this method has been applied to analyse linear data (Qiang et al., 2014); however, it has rarely been used to analyse movement data.

Chapter 2 presents the results of an analysis of football movement data using the CTM. The research focuses on the exploration of a match performance, which is considered an important issue in the domain of sports analytics. The results demonstrate that the CTM is indeed useful in exploring match performances, discovering insightful information, and providing potential suggestions to sports professionals (e.g., coaches) because compared with current methods that are frequently used in football match analysis, the CTM can visualise continuous changes of information (e.g., speed, ball possession, territorial advantage) at any temporal scale from the starting timestamp until the ending timestamp in one single figure. In this way, sports professionals (e.g., coaches) can gain an overall and detailed grasp of changes in player data. Therefore, the proposed method increases the added value compared with traditional methods and might have more options for further tactical arrangements, which represents one of the strengths of applying the CTM to the analysis of football movement data. A second strength is that in addition to representing information that can be derived by spatio-temporal data, the CTM also has the ability to visualise semantic information, which might be highly interesting to sports professionals. Important semantic information, such as ‘shot on target’ and ‘goal’ (which are two important factors that determine whether a team can win or not), can be clearly visualised in the CTM diagrams. With this information, sports professionals can obtain a clear overview of the desired semantic information and more conveniently explore the underlying reasons to identify potential suggestions for improving training or tactical arrangements. Because the abilities of the CTM can be enhanced by the support of map algebra operations (e.g., summation, subtraction, maximum, minimum, mean, median), this method has strong extensibility and provides various functions to achieve many simple or advanced analyses or investigations. Hence, potential future research into the CTM could focus on incorporating the CTM into an interactive graphical user interface, which could augment the usability of the CTM for sports analytics, and incorporating the CTM with a heat map, which is considered as a useful and popular visualisation tool in team sports analysis. Thus, the ability of the heat map can be enhanced to address temporal information. In all, the CTM is regarded as a novel tool for analysing movement data. We are convinced that more extensive applications with respect to the CTM can emerge in the domain of movement data analysis in the future.

Next, we conducted research on knowledge discovery in movement data. Because data mining has already attracted considerable attention in this data-rich era, it has been considered a key technique for knowledge discovery in various data (Han & Kamber, 2006). Thus, data mining techniques can be used to identify useful knowledge in movement data as well. Useful knowledge includes movement patterns and important time intervals, which are two main foci of this thesis.

Chapter 5 mainly focuses on movement pattern discovery in movement data. Various types of movement patterns are observed in movement data, such as flock patterns, convoy patterns, leadership patterns, moving clusters, and crews. In this thesis, we are particularly interested in moving flock patterns. Definitions of moving flock patterns are still lacking, and few common taxonomies of moving flock patterns are available to divide moving flock patterns into various categories. Therefore, as one of the main contributions in this chapter, we proposed an improved definition of 'moving flock' in Chapter 5 based on the definition proposed in Wachowicz et al. (2011). Compared with the original definition of moving flock, the improved definition can identify moving flock patterns in movement data more accurately. Another main contribution is the proposed taxonomy of moving flock patterns, which is used to derive a number of different types of moving flock patterns. The desired types of moving flock patterns can be derived according to specific demands. Specifically, we are interested in eight types of moving flock patterns. To automatically identify these types of moving flock patterns in movement data, we developed a Reeb graph-based approach, which is considered the third main contribution. The proposed approach was applied to football movement data obtained from a real football match, and the results indicate that the proposed approach is capable of discovering the desired moving flock patterns. In this approach, the Reeb graph is employed as a novel tool to model movement data as a graph, and based on the generated graph, corresponding algorithms are designed to discover the moving flock patterns. Although the Reeb graph has been extensively used in many fields, such as shape analysis (Biasotti et al., 2008; Chen et al., 2013) and scientific data visualisation (Fomenko & Kunii, 1997; Edelsbrunner & Harer, 2010), it has rarely been adopted in GIScience. In the future, other types of movement patterns might be discovered by developing corresponding Reeb graph-based approaches. If so, Reeb graphs can be adopted as an interesting tool for analysing movement data or even extended to analyse other types of spatio-temporal data. In addition, this approach consists of four parameters. Although we have investigated the influences of each parameter on the corresponding results and the findings demonstrate that the results are meaningful under the suggested/chosen parameter combinations, the selection of optimal parameter values is still an open and difficult question. Considerable research still needs to be conducted either to investigate how the parameters influence the results to determine more optimal parameter values or to improve the approach by involving less parameters.

The work presented in Chapter 3 can be considered within the scope of knowledge discovery in movement data. The main contribution of Chapter 3 is the development of a cross-scale oriented sequence analysis approach to discovering knowledge in movement data. Scale is considered an important factor in this chapter in the development of the approach because scale is a common problematic issue in many disciplines, particularly when space and time are important components (e.g., GIScience). Correspondingly, scale can be classified as the spatial scale, temporal scale and spatio-temporal scale. The scale appears even more important, especially when it has been considered as the fifth dimension in 5D data modelling (van Oosterom & Stoter, 2010). Given the importance of scale, we primarily aim to explore the effects of scale in the analysis of movement data. In this chapter, we mainly focus on the temporal scale. The effects of temporal scale on the changes of movements of objects are explored by considering the values of motion attributes (e.g., speed) across a large number of temporal scales. The results reveal how specific temporal scales affect the movement changes of objects. Another important contribution of this chapter is the development of an approach for identifying the time intervals during which important events might occur because such data are very important to specific people. For example, the time intervals during which traffic congestion might occur are important to traffic planning and management-related specialists. In sports, these data are also important. For example, in football, specific actions, such as a goal and a shot, are of interest to coaches because they can be used to analyse the performance of players or the whole team. Thus, the time intervals during which these events occur are important. In this approach, temporal scales were involved when detecting such time intervals, which is novel compared with approaches that do not consider temporal scales. The results demonstrate that the approach is capable of detecting the time intervals during which important events occur and can generate superior results compared with methods that only consider one temporal scale. In this chapter, only speed is used as an example when exploring the changes of motion attributes, and other attributes, such as distance and motion azimuth, might be explored using the proposed approach in the future. Only the temporal scale is considered in this chapter and this thesis. In addition to the temporal scale, the spatial scale and spatio-temporal scale are also important. In future research, either the spatial scale or the spatio-temporal scale can be considered. Combining the spatial or spatio-temporal scale with the research outcomes with respect to the temporal scale in this thesis can generate clear insights into the influence of scale on the analysis of movement data. This approach is also quite extensive; thus, it can also be applied to movement data in other domains.

The last research focus in this thesis is the dynamic interactions in movement data. In this thesis, we only considered the interactions between/among moving objects themselves. Furthermore, interactions can be classified as static interactions and dynamic interactions (Doncaster, 1990). Similar to spatio-temporal data, dynamic interactions are defined based on both the spatial and temporal components, whereas static interactions are purely described by the spatial properties (Miller, 2015). In this work, we are more interested in the dynamic interactions. The existing research on dynamic interactions mainly focuses on either comparing or evaluating existing interaction methods based on different datasets. Dynamic interactions are primarily explored between two moving objects, and previous research on dynamic interactions has generally been performed at a single temporal scale. Hence, an exploration of dynamic interactions among multiple moving objects and at multiple temporal scales is needed. To our knowledge, previous studies have not focused on exploring the importance of each object and identifying the objects that play relatively important roles in maintaining specific types of interaction patterns.

Accordingly, Chapter 4 presents a hybrid approach that combines the Multi-Temporal Scale Spatio-Temporal Network (MTSSTN) and the CTM to meet the aforementioned demands. This approach represents the key contribution of Chapter 4, and it is applied to analyse football movement data. The results show that the proposed approach is capable of exploring various interactions either between two players or among multiple players at any temporal scale, and it is also useful in evaluating the importance of each player and identifying the most important players. The most obvious characteristics of the proposed approach are its superiority in analysing dynamic interactions at multiple temporal scales compared with previous methods and ability to explore the dynamic interactions among multiple (usually at least three) objects. Another distinctive characteristic proposed in this work is the MTSSTN, which is a rather novel type of network that is more advanced than the Spatio-Temporal Network (STN) recently developed by Williams & Musolesi (2016) because it can address multi-temporal scale related issues while the STN cannot. In this chapter, the interactions were explored based on the corresponding interaction patterns generated in terms of the Relative Trajectory Calculus (RTC) (Van de Weghe, 2004). Although the results demonstrate the meaning and effectiveness of the derived interaction patterns, the interaction patterns still appear simple. In future research, more sophisticated interaction patterns might be derived using other potential methods to enhance the meanings of the interaction patterns. Optional methods include the Qualitative Trajectory Calculus (QTC) (Van de Weghe et al., 2004; Van de Weghe et al., 2005; Van de Weghe et al., 2006), Dynamic Interaction (DI) (Long & Nelson, 2013), and RElative MOtion (REMO) (Laube et al., 2005). In this way, more meaningful and insightful information on dynamic interactions in movement data might be revealed.

6.2 Conclusions

In recent decades, a dramatic improvement of positioning technologies has been observed, and it has led to the generation of massive volumes of tracked data on virtually any object that moves, which are called ‘movement data’ in this thesis. Thus, many new methods for managing these data have been developed and applied to a multitude of application domains. However, new approaches for analysing such data and extensions of approaches to novel applications are required. Motivated by these demands, this thesis mainly aims to develop new approaches for the analysis of movement data and extend the developed approaches to a relatively novel application domain: sports. Thus, this thesis contributes to both methodological and application-oriented research. Four general research questions are posed in this thesis. Based on the research questions, four chapters are subsequently presented, and each can answer the research questions in whole or in part. The answers provided to resolve these questions are considered the main contributions of this thesis.

This thesis presents an original research effort on the use of the CTM to analyse football movement data obtained from a real and entire football match. Although the CTM has already been proposed by Qiang et al. (2014), it has not been used broadly and thus has the potential to be extended significantly to a large number of domains in the future. Moreover, the functionalities of the CTM could be extended to a considerable degree. Hence, in this chapter, we contribute to developing the functionalities of the CTM and the initial applications of the CTM in football. The applications are achieved via visual explorations of a match performance based on the obtained football movement data. In general, the performance of players and teams is explored according to the CTM diagrams of various motion attributes. Specifically, the motion attributes include several basic motion attributes and one more complex motion attribute, which builds upon multiple basic attributes. As for the basic motion attributes, further operations are executed based on map algebra operators to identify more insightful information. The results demonstrate that the CTM can indeed be used to serve the domain of sports analytics (e.g., exploring a match performance) and thus can be used by sports professionals to obtain distinct information that cannot be obtained by traditional approaches in sports analytics or data that are difficult to obtain. Because the main focus of this thesis is developing new approaches for analysing movement data, this application for sports analysis (i.e., football) can be considered a first step. In the future, the proposed method can be combined with more detailed interpretations by domain specialists. We are convinced that this research can be extended significantly to better serve the domain of sports analytics in the future.

This thesis contributes to knowledge discovery in movement data by developing a cross-scale oriented sequence analysis approach. The key aspects of this approach are the four different types of sequences that are constructed based on the CTM, and each of the sequences is considered based on varying temporal scales. Based on the four types of sequences, two specific aims are achieved: investigating the changes of motion attributes across different temporal scales (based on the first type of sequences) and detecting the time intervals during which important events might occur (based on the remaining three types of sequences). The results show that more abundant information can be gained by investigating the changes of motion attributes across different temporal scales, and the time intervals detected by the proposed approach are more accurate than approaches that do not consider multiple temporal scales. This finding demonstrates the usefulness and effectiveness of the proposed approach. A hybrid approach that combines the MTSSTN and the CTM was proposed in this thesis to explore the dynamic interactions in movement data. Specifically, the dynamic interactions are explored based on specific interaction patterns derived by the RTC based on an exploration of the interaction intensities between two individuals or among multiple individuals as well as on the importance of each individual and identifying the most important individuals for maintaining each type of interaction pattern. The proposed approach is validated with the football movement data.

The results demonstrate that the proposed approach is useful for exploring dynamic interactions in movement data. We believe that this research can play specific roles in the relatively new domain of dynamic interactions in movement data in the future.

This thesis also contributes to the discovery of movement patterns in movement data. Specifically, the primary objective of this thesis is to discover moving flock patterns, which is of particular interested. The main contributions in this chapter are threefold: first, an improved definition of 'moving flock' is proposed, and it can be used to more accurately characterise moving flock patterns; second, a taxonomy of moving flock patterns is developed, and it can be used to derive a large amount of moving flock patterns; third, a Reeb graph-based approach is developed. Among the large amount of moving flock patterns, we are particularly interested in eight types, and the Reeb graph-based approach is developed to identify the desired moving flock patterns. The football movement data are used to validate the proposed approach. The results show the effectiveness of the proposed approach in discovering moving flock patterns and the potential usefulness of the proposed approach in providing insightful information to domain experts. The results indicate that the approach has much potential for use in other application domains as well.

References

Benkert, M., Gudmundsson, J., Hübner, F., & Wolle, T. (2008). Reporting flock patterns. *Computational Geometry, 41*(3), 111-125.

Biasotti, S., Giorgi, D., Spagnuolo, M., & Falcidieno, B. (2008). Reeb graphs for shape analysis and applications. *Theoretical Computer Science, 392*, 5-22.

Chen, F., Obermaier, H., Hagen, H., Hamann, B., Tierny, J., & Pascucci, V. (2013). Topology analysis of time-dependent multi-fluid data using the Reeb graph. *Computer Aided Geometric Design, 30*(6), 557-566.

Doncaster, C. P. (1990). Non-parametric estimates of interaction from radio-tracking data. *Journal of Theoretical Biology, 143*(4), 431-443.

Edelsbrunner, H., & Harer, J. L. (2010). *Computational topology: an introduction*. American Mathematical Society.

Fomenko, A., & Kunii, T. (1997). *Topological methods for visualization*. Springer, Tokyo, Japan.

Fort, M., Sellarès, J. A., & Valladares, N. (2014). A parallel GPU-based approach for reporting flock patterns. *International Journal of Geographical Information Science*, *28*(9), 1877-1903.

Gudmundsson, J., van Kreveld, M., & Speckmann, B. (2007). Efficient detection of patterns in 2D trajectories of moving points. *Geoinformatica*, *11*(2), 195-215.

Han, J., & Kamber, M. (2006). *Data mining: concepts and techniques*. Morgan kaufmann.

Kalnis, P., Mamoulis, N., & Bakiras, S. (2005). On discovering moving clusters in spatio-temporal data. In *International Symposium on Spatial and Temporal Databases* (pp. 364-381).

Laube, P., Imfeld, S., & Weibel, R. (2005). Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science*, *19*(6), 639-668.

Loglisci, C. (2017). Using interactions and dynamics for mining groups of moving objects from trajectory data. *International Journal of Geographical Information Science*, 1-33.

Long, J. A., & Nelson, T. A. (2013). Measuring dynamic interaction in movement data. *Transactions in GIS*, *17*(1), 62-77.

Miller, J. A. (2015). Towards a better understanding of dynamic interaction metrics for wildlife: a null model approach. *Transactions in GIS*, *19*(3), 342-361.

Qiang, Y., Chavoshi, S. H., Logghe, S., De Maeyer, P., & Van de Weghe, N. (2014). Multi-scale analysis of linear data in a two-dimensional space. *Information Visualization*, *13*(3), 248-265.

Solera, F., Calderara, S., & Cucchiara, R. (2015). Learning to identify leaders in crowd. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 43-48).

Turdukulov, U., Calderon Romero, A. O., Huisman, O., & Retsios, V. (2014). Visual mining of moving flock patterns in large spatio-temporal data sets using a frequent pattern approach. *International Journal of Geographical Information Science*, *28*(10), 2013-2029.

Van de Weghe, N. (2004). *Representing and reasoning about moving objects: a qualitative approach*. Ghent University

Van de Weghe, N., Cohn, A. G., De Maeyer, P., & Witlox, F. (2005). Representing moving objects in computer-based expert systems: the overtake event example. *Expert Systems with Applications*, *29*(4), 977-983.

Van de Weghe, N., Cohn, A. G., De Tre, G., & De Maeyer, P. (2006). A qualitative trajectory calculus as a basis for representing moving objects in geographical information systems. *Control and Cybernetics*, *35*(1), 97-119.

van Oosterom, P., & Stoter, J. (2010). 5D data modelling: full integration of 2D/3D space, time and scale dimensions. *Lecture Notes in Computer Science*, *6292*, 310-324.

Wachowicz, M., Ong, R., Renso, C., & Nanni, M. (2011). Finding moving flock patterns among pedestrians through collective coherence. *International Journal of Geographicl Information Science*, *25*(11), 1849-1864.

Williams, M. J., & Musolesi, M. (2016). Spatio-temporal networks: reachability, centrality and robustness. *Royal Society open science*, *3*(6), 160196.

Yeoman, J., & Duckham, M. (2016). Decentralized detection and monitoring of convoy patterns. *International Journal of Geographical Information Science*, *30*(5), 993-1011.

Summary

With the development of location-aware technologies, massive volumes of tracked data on virtually any object that moves have emerged, which has led to a proliferation of movement data. Thus, new approaches for analysing such data are greatly needed. In addition, the ability to extend available approaches to new domains is also in demand. Motivated by these developments, this thesis mainly focuses on developing new approaches for the analysis of movement data. As a first step, the approaches developed in this thesis are used in a relatively novel application domain: sports. Specifically, football movement data obtained from a real and entire football match are adopted to support the methodological and sports-oriented research in this thesis. In this section, the contributions with respect to the major research questions proposed in Chapter 1 are summarised.

RQ 1: Can the CTM bring added value to the analysis of movement data?

The Continuous Triangular Model (CTM) has two distinct characteristics in representing temporal information: visualisation and multiple scales. Hence, in this thesis, to explore the added value of the current research in the analysis of movement data, the CTM is used and developed as a tool to analyse movement data, either from a visualisation perspective or from a multi-scale perspective. This research question is addressed in Chapters 2, 3 and 4.

Chapter 2 presents an exhaustive application of the CTM in analysing movement data using football movement data obtained from a real and entire football match as an example. In this study, the functionalities of the CTM are extended and presented based on the movement data corresponding to an entire football match (from the beginning to the end), which to our knowledge is a first use of the CTM for such a large dataset. The performance of both the players and the whole team can be explored and evaluated according to the CTM diagrams of corresponding motion attributes, either from a visualisation perspective or a multi-scale perspective. The results demonstrate that the CTM is indeed useful in exploring a match performance and discovering insightful information. In Chapter 4, the main focus is on the exploration of dynamic interactions in movement data by developing a hybrid approach. The hybrid approach combines the Multi-Temporal Scale Spatio-

Temporal Network (MTSSTN) and the CTM. The CTM is used as a reference to construct a MTSSTN with a CTM-like shape. Based on the constructed MTSSTNs, the interaction intensity measures between/among objects can be calculated and subsequently visualised using the CTM diagrams and the importance of individual objects can be determined. According to the corresponding CTM diagrams, the interaction intensities between/among objects, the importance of each individual, and the most important individuals in each interaction pattern can also be explored from a visualisation perspective or a multi-scale perspective. The results demonstrate the effectiveness of the proposed approach in the application of football. The research presented in both Chapters 2 and 4 show the applicability of the CTM in analysing movement data.

Chapter 3 mainly focuses on the multi-scale characteristics of the CTM. In this chapter, a cross-scale oriented sequence analysis approach is proposed for knowledge discovery in movement data. The main contribution is the construction of the four types of sequences on the basis of the CTM. For the first type, the discrete points in the same sequence all have the same temporal scales. For the remaining three types, the discrete points in the same sequence all have different temporal scales. Based on the first type of sequences, the movements of objects across different temporal scales can be characterised according to the changes of corresponding motion attributes. Based on the combination of the remaining three types of sequences, the time intervals during which important events might occur can be detected. The application of the proposed approach in the football movement data shows its effectiveness and advantages in the analysis of movement data.

RQ 2: What interesting information can be discovered in football movement data?

Knowledge discovery in various types of data is a topic that has attracted the attention of researchers from different domains for a long time. Similar to other types of data, knowledge discovery in movement data can also be achieved by data mining techniques. As a novel type of movement data, various sports movement data have recently become available. Hence, interesting information could be discovered in such data using movement data mining techniques. To explore this aspect, two new data mining approaches for movement data are developed and presented in Chapter 3 and Chapter 5. The approaches are applied to football movement data to discover interesting information. Based on the approach developed in Chapter 3, the changes of movements across different temporal scales can be characterised. In addition, the time intervals during which important events might occur can be detected. These findings are relatively novel compared with those of traditional methods for performing football analyses. Chapter 5 develops a Reeb graph-based approach to automatically discovering various types of moving flock patterns. The members in a moving flock pattern can be simply understood by the characteristic 'keep spatially close and move together for a specific duration'; thus, such players are rather interesting to coaches because they might show close interactions with each other. Based on the approach developed in Chapter 5, various groups of such players can be discovered and additional insightful information can be provided to coaches. The results in both chapters show that interesting information can indeed be discovered in football movement data using the methods for analysing movement data; thus, the proposed approaches for the analysis of movement data have the potential for use in various sports domains.

RQ 3: Can added values be provided if multiple (temporal) scales are considered when analysing movement data?

Scale is an important inner attribute of almost any type of movement data. Scale can be classified at the spatial scale, temporal scale and spatio-temporal scale. In this thesis, we mainly focus on the temporal scale. Chapters 3 and 4 both consider the temporal scale when developing respective approaches for the analysis of movement data. In Chapter 3, a cross-scale-oriented sequence analysis approach is proposed. According to this approach, the changes of movement across different temporal scales are characterised. Such changes are not easily characterised without considering temporal scales. In addition, the time intervals during which important events might have occured can be detected with higher accuracy using the proposed approach than methods that only consider one temporal scale. In Chapter 4, a hybrid approach combining the MTSSTN and the CTM is proposed. When constructing the MTSSTN, the temporal scales are considered. Based on the MTSSTN, the information at multiple temporal scales can be calculated and then visualised using the CTM. According to the CTM diagrams, much more abundant information can be acquired. Hence, added value can indeed be provided when multiple (temporal) scales are considered when analysing movement data.

RQ 4: What efforts can be contributed to the relatively new research topic of dynamic interactions in movement data?

The research on dynamic interactions in movement data is relatively new compared with other research topics with respect to the analysis of movement data. Therefore, many potential efforts can be devoted to this topic. In this thesis, one research effort is contributed to this topic and addressed in Chapter 4. In this chapter, a hybrid approach combining the MTSSTN and the CTM is proposed. The proposed approach is capable of performing a quantitative exploration of the interactions between two individuals as well as the interactions among multiple individuals, which have seldom been studied, and this capability represents one of the main contributions of this thesis. Another contribution is the ability to explore the importance of each individual in the interactions and identify the most important individuals in each interaction. By applying the proposed approach to football movement data, interesting results have been obtained, which shows the effectiveness of the proposed approach. Thus, the approach proposed in this thesis can contribute to revealing the dynamic interactions in movement data and promote the development of this research topic in the future.

Samenvatting

De explosie van locatiebewuste technologieën zorgt voor uitgebreide datasets van bewegingsgegevens. Nieuwe methodes om zulke gegevens te analyseren zijn noodzakelijk. Geïnspireerd door deze ontwikkelingen, richt deze thesis zich voornamelijk op de ontwikkeling van nieuwe methodes met betrekking tot de analyse van bewegingsgegevens. De benaderingen ontwikkeld in deze thesis worden in eerste instantie toegepast op een relatief nieuw domein: sport. Meer specifiek worden in deze thesis de voetbalverplaatsingsgegevens (verkregen van een echte en volledige voetbalwedstrijd) benut om het methodologisch en sportgeoriënteerd onderzoek te ondersteunen. In deze sectie worden de bijdragen aangaande de belangrijkste voorgestelde onderzoeksvragen gepresenteerd.

RQ 1: Kan het CTM een meerwaarde bieden bij de analyse van de bewegingsgegevens ?

Het Continu Triangulair Model (CTM) heeft twee uitgesproken eigenschappen in het voorstellen van tijdelijke informatie: visualisatie en meerschaligheid. Daarom werd in deze thesis het CTM gebruikt om de meerwaarde van het huidige onderzoek in bewegingsgegevensanalyse te bekijken. Deze onderzoeksvraag werd behandeld in de hoofdstukken 2, 3 en 4.

Hoofdstuk 2 stelt een grondige CTM-toepassing voor in het ontleden van voetbalbewegingsgegevens die verkregen werden uit een echte en volledige voetbalwedstrijd. In deze studie worden de functionaliteiten van het CTM voorgesteld en uitgebreid, op basis van de verplaatsingsgegevens die overeenstemmen met een volledige voetbalmatch (van begin tot einde). Dit is - voor zover wij op de hoogte zijn - de eerste CTM-toepassing op een dergelijk uitgebreide dataset. De prestatie van zowel de spelers als het hele team kan onderzocht en geëvalueerd worden volgens de CTM-representaties van overeenstemmende bewegingsattributen, zowel vanuit visualisatieperspectief als vanuit meerschalig oogpunt. De resultaten geven aan dat het CTM inderdaad nuttig is in het ontdekken van wedstrijdprestaties. In hoofdstuk 4 ligt de nadruk op het vinden van dynamische interacties in de bewegingsgegevens door het ontwikkelen van een hybride aanpak. Deze aanpak combineert het *Multi-Temporal Scale Spatio-Temporal Network*

(MTSSTN) en het CTM. Het CTM wordt gebruikt als referentie om een MTSSTN met een CTM-achtige vorm te construeren. Gebaseerd op de gebouwde MTSSTN’s kunnen de intensiteiten van de interacties tussen objecten berekend worden. Vervolgens kunnen deze gevisualiseerd worden door middel van de CTM-representaties en kan het belang van individuele objecten bepaald worden. Volgens de overeenstemmende CTM-representaties kunnen de intensiteiten van de interacties tussen objecten, het belang van ieder individu en de meest belangrijke personen in elk interactiepatroon onderzocht worden vanuit een visualisatieperspectief of een meerschalig oogpunt. De resultaten geven de effectiviteit weer van de voorgestelde aanpak in de voetbalapplicatie. Het gepresenteerde onderzoek in de hoofdstukken 2 en 4 toont de CTM-toepasbaarheid aan in de analyse van bewegingsgegevens.

Hoofdstuk 3 richt zich hoofdzakelijk op de meerschalige eigenschappen van het CTM. In dit hoofdstuk wordt een zogenaamde schaaloverschrijdende georiënteerde sequentie-analyse aanpak voorgesteld om kennis te verkrijgen uit de verplaatsingsgegevens. De belangrijkste bijdrage is de creatie van vier soorten sequenties gebaseerd op het CTM. Bij de eerste soort hebben de afzonderlijke punten in dezelfde sequentie allemaal dezelfde temporele schaal. Bij de resterende drie soorten hebben de afzonderlijke punten in dezelfde sequentie allemaal een verschillende temporele schaal. Gebaseerd op de eerste soort kunnen de objectbewegingen over verschillende temporele schalen getypeerd worden volgens de wijzigingen van de corresponderende attribuutbewegingen. Steunend op de combinatie van de drie resterende soorten sequenties kunnen tijdsintervallen (gedurende dewelke belangrijke gebeurtenissen kunnen plaatsvinden) gedetecteerd worden. De toepassing van de voorgestelde methode op deze voetbalbewegingsgegevens bewijst de effectiviteit en de voordelen van het ontleden van verplaatsingsgegevens.

RQ 2: Welke interessante informatie kan verkregen worden aan de hand van voetbalbewegingsgegevens ?

Het achterhalen van kennis in verschillende types van gegevens is een onderwerp dat sinds geruime tijd de aandacht trekt van onderzoekers uit verschillende domeinen. Vergelijkbaar met andere soorten gegevens kan informatie verkregen worden uit bewegingsgegevens door het gebruik van *data mining* methodes. Om dit aspect te onderzoeken worden twee nieuwe data mining methodes (voor bewegingsgegevens) ontwikkeld en voorgesteld in hoofdstuk 3 en 5. Deze methodes worden toegepast op voetbalverplaatsingsgegevens. Gebaseerd op de ontwikkelde methode in hoofdstuk 3 kunnen de bewegingsveranderingen over verschillende temporele schalen gekarakterisererd worden. Bovendien kunnen de tijdsintervallen - gedurende dewelke belangrijke gebeurtenissen kunnen plaatsvinden - achterhaald worden. Deze bevindingen zijn vernieuwend vergeleken met deze die afkomstig zijn van traditionele methodes voor het uitvoeren van voetbalanalyses. Hoofdstuk 5 behandelt de op de Reeb-grafiek geënte aanpak om verscheidene types voortbewegende *moving flock* patronen automatisch op te sporen. Leden van een *moving flock* zijn objecten (bv. spelers) die samen bewegen op een korte afstand van elkaar gedurende een bepaalde tijd; zulke spelers kunnen dus interessant zijn voor coaches omdat ze nauwe interacties met elkaar kunnen vertonen. Op basis van de ontwikkelde aanpak beschreven in hoofdstuk 5 kunnen verschillende groepen bestaande uit dergelijke spelers gedetecteerd worden en kan relevante informatie aan coaches bezorgd worden. De resultaten in beide hoofdstukken tonen aan dat interessante informatie inderdaad verkregen kan worden. De voorgestelde aanpak heeft dus potentieel voor toepassingen in verschillende sportdomeinen.

RQ 3: Kan er een meerwaarde geleverd worden indien er meerdere (temporele) schalen in overweging genomen worden bij het ontleden van de bewegingsgegevens ?

De schaal is een belangrijke interne kenmerkende eigenschap van bijna elk type verplaatsingsgegevens. Schaal kan ruimtelijk, temporeel en tijdruimtelijk zijn. In deze thesis focussen we ons voornamelijk op de temporele schaal. Hoofdstukken 3 en 4 bestuderen de temporele schaal bij het ontwikkelen van methodes voor de analyse van bewegingsgegevens. In hoofdstuk 3 wordt een schaaloverschrijdende volgorde-analyse (*crossscale-oriented sequence analysis*) voorgesteld. Volgens deze aanpak worden de bewegingswijzigingen op verschillende temporele schalen gekarakteriseerd. Zulke veranderingen worden niet eenvoudig herkend zonder verschillende temporele schalen in rekening te nemen. Bovendien kunnen de tijdsintervallen (gedurende dewelke belangrijke gebeurenissen zouden kunnen plaatsgevonden hebben) gedetecteerd worden met een hogere nauwkeurigheid door middel van de voorgestelde aanpak dan met methodes die enkel één temporele schaal in rekening nemen. In hoofdstuk 4 wordt een hybride aanpak (die het MTSSTN en het CTM combineert) besproken. Bij het construeren van het MTSSTN worden de verschillende temporele schalen behandeld. Gebaseerd op het MTSSTN kan de informatie berekend en daarna gevisualiseerd worden op meerdere temporele schalen door middel van het CTM. Via de CTM-representaties kan veel informatie verworven worden. Er kan dus een meerwaarde verkregen door meerdere temporele schalen te gebruiken tijdens het analyseren van bewegingsgegevens.

RQ 4: Welke inspanningen kunnen bijdragen tot het relatief nieuwe onderzoekstopic over dynamische interacties in bewegingsgegevens ?

Het onderzoek van de dynamische interacties van verplaatsingsgegevens is nieuw in vergelijking met andere onderzoeksthema's met betrekking tot de analyse van bewegingsgegevens. In hoofdstuk 4 wordt hier dieper op ingegaan. In dit hoofdstuk wordt een hybride methode (die het MTSSTN en het CTM combineert) voorgesteld. De voorgestelde aanpak onderzoekt op een kwantitatieve manier de interacties tussen meerdere personen; wat zelden bestudeerd werd. Dit vormt een van de voornaamste bijdragen van deze thesis. Een andere belangrijk bijdrage is om de meest belangrijke personen in elke interactie te identificeren. Door de voorgestelde aanpak op voetbalbewegingsgegevens toe te passen werden interessante resultaten verkregen die de effectiviteit van de voorgestelde methode aantonen. De voorgestelde aanpak in deze thesis kan dus bijdragen tot de analyse van de dynamische interacties in bewegingsgegevens en kan de ontwikkeling van dit onderzoeksthema in de toekomst bevorderen.