Chapter 1

Introduction

1.1 Context and Motivation

In recent years, improvements in technology have generated huge volume of data that require analysis. Interest in analysis of movement data has grown manifold. Applications, with significant emphasis on such analysis, have emerged. Works relating to movement data analysis can be found in two different areas. One area is Geographic Information Science (GIScience). Here, analysis of movement of objects in a wide coverage area is performed. Spatial entities are typically human beings, animals, birds, vehicles etc. Movements of such entities over a geographical space are analysed for gathering meaningful information. This meaningful information is generally in the form of patterns of movement exhibited by objects. Movement patterns or motion patterns include any recognizable spatial or temporal regularity or any interesting relationship in a set of movement data [1]. This statement mentions two characteristics of a movement pattern. Firstly, in motion patterns, we look for some regularity of movement in terms of space, time or space and time. Secondly, a pattern can also express some interesting relationship in the movements of the participating entities, though no *regularity* can be seen. We would like to mention here that we are using the terms *movement pattern* and *motion pattern* synonymously. Movement pattern recognition is useful in understanding social behaviour of animals, group behaviours, migration patterns, pattern of movement of people using mobile phones, tourist movement patterns etc. These data are typically in the form of coordinate points in some reference system annotated by some temporal information about the points.

Another area is about extracting high level semantics from video data. Improvements in digital technology have resulted in availability of huge volume of video data. With advancements in object detection and tracking, it has been possible to perform such analysis with more precision. Recognition of activities from video has become an important application in computer vision. In computer vision literature, the term *action* refers to simple motion patterns usually executed by a single person and lasting for a short interval of time. An *activity* involves complex sequence of actions executed by several humans who can have interactions and an *activity* lasts for a longer duration of time [3]. Motion pattern recognition is the principle part of activity recognition. There are many areas where activity recognition becomes a focus of interest. Visual surveillance is one such application where one would like to recognise abnormal activities like suspicious movement of people in public places, unusual grouping of people etc. Traffic monitoring is another area where we want to detect cases of traffic rule violation; movements that can potentially cause accidents, congestion etc. In sports analysis, for example in a football game, we may be interested to know different patterns that indicate strategies of attack or defense. In other words, the point of interest is the semantics of the game. A patient monitoring system may analyse motion patterns of patients in a room.

Movement data are treated mostly in two dimensional form. In video analysis, 3D world is projected onto a 2D image [4]. In the case of GIScience, these data are in the form of coordinate points in two dimensional form indicating the positions of the objects over a time interval [1]. Such numerical data about movement do not convey much information to a human being. It is the interesting relationship in data that the human mind is eager to explore. In order to explore such interesting relationships, movement data are generally represented along multiple dimensions. These are the dimensions humans use for their commonsense understanding of movement. Typical dimensions include speed, direction, distance, spatial orientation etc. These dimensions can be called as *elements* or *parameters* of a movement pattern. It is possible to express patterns of movement by specifying precise quantitative values for movement parameters. For example, we can say that two cars are moving together, the distance between them varies from 2m to 3m and speed of movement varies from 5 km/hr to 10 km/hr. Here, we have provided movement data for two movement parameters using precise quantitative data. Perhaps, from the point of commonsense understanding, the same thing can be stated as: two cars are moving slowly at close distance.

A *qualitative*, rather than *quantitative*, approach may be more adequate for representation of spatio-temporal human cognition [5], [6]. Qualitative formalisms have gained acceptance as a method of abstracting away from metrical details. Qualitative reasoning represents continuous properties of a system by a discrete set of symbols. Knowledge is represented using qualitative relationships among entities or qualitative categories of numerical values. A qualitative distinction is introduced only when it is relevant to the problem [7]. An example of qualitative category is the set $\{+, -, 0\}$. These three symbols in the set can be used to represent real numbers. On the other hand, when knowledge is represented in the form of relationships between objects, we arrive at a set of binary qualitative relations. For reasoning tasks, this set of binary qualitative relations needs to be Jointly Exhaustive and Pairwise Disjoint (JEPD). Qualitative Spatial Reasoning (QSR) is the field of knowledge representation and reasoning that uses this type of qualitative approach. In QSR, qualitative formalisms have been proposed for spatial aspects like topology, orientation, direction, distance etc. [8]. For recognition of a movement pattern, QSR can be applied for creating high level qualitative abstraction because of the qualitative nature of our common sense reasoning. Works on modelling movement parameters qualitatively are reported [9], [10]. Though activity recognition involves lot of video processing tasks at the low level, interactions at the high level are modelled using qualitative relationships [11], [12].

Movement patterns possess a hierarchical structure. This hierarchy is implicit in our everyday description of movement. As an example, a statement

Figure 1.1: Movement Pattern: A Hierarchical Perspective

about movement of four objects can be like: *Two persons were walking along the road and approaching each other from opposite directions; at the same time two others were crossing the road from opposite directions. These two persons had crossed the road before the first two arrived.* This movement pattern is illustrated in Figure 1.1. This pattern can be broken down into two sub-patterns. One is the *approach_from_opposite* pattern between two persons from opposite directions and the other is the *cross_from_opposite* of two persons. There is a temporal relationship between these sub-patterns. The *cross_from_opposite* sub-pattern completes before the *approach_from_opposite* sub- pattern. This sort of hierarchical description is possible because humans already have an intuitive idea about what constitutes a *cross_from_opposite* or *approach_from_opposite* pattern. It is true that for the movement pattern shown in Figure 1.1 there can be different cognitive interpretations. Let the persons at Left, Right, Top and Bottom be designated as L, R, T and B. We have taken into account the movement of L with respect to R and T with respect to B. In another cognitive interpretation, we can interpret that with respect to B, L crosses in front of him from left to right and with respect to T, R crosses in front of him from right to left. This video was demonstrated to a group of twenty persons and they were requested to give a cognitive interpretation. This cognitive analysis is presented in Appendix E. Selection of the object with respect to which we want to describe the motion pattern is a subjective issue; but it affects learning, representation and recognition of such patterns. This issue is discussed in chapter 5 and also in chapter 6.

Hierarchical structure of processes is represented and recognised by a field of pattern recognition known as syntactic pattern recognition. Syntactic pattern recognition uses formal grammars or relational graph matching for recognition of patterns. In grammar based recognition, a large number of motion patterns can be represented in a concise form using formal grammars. Moreover, efficient algorithms may be used for the recognition task. Works on use of grammar based syntactic approaches for recognising activities from video samples have been reported in the literature [13], [14].

Motivation of this work is to have a mechanism for hierarchical qualitative representation and recognition of movement patterns. Hierarchical representation is significant because it allows us to construct patterns from known primitives. For an application domain, standard patterns can be predefined and these standard patterns can be used to represent patterns at the next level of abstraction. For example, in the domain of traffic analysis, standard patterns like *overtake*, *block*, *follow* etc. can be predefined. This type of exploration of the internal structure of motion pattern will convey more meaningful information. At the same time, this representation will be more relevant for commonsense understanding. Moreover, qualitative information may be more available and less expensive to obtain than quantitative information [15]. Qualitative representations may also prove to be useful for handling incomplete information [16].

1.2 Objectives and Significance

The objective of this work is *to provide a framework that integrates qualitative spatial reasoning with grammar based syntactic pattern recognition for representation and recognition of motion patterns.* This work is significant in activity recognition in video processing. This work can find use in video analysis applications such as security and surveillance, content based video analysis, animation and synthesis etc. In GIScience, classification of movement patterns have been proposed over the years. Such patterns like *flock*, *leadership*, *convergence*, *colocation*, *trend-setting*, *moving cluster* etc. are important for a study of migration patterns, group behaviour, social behaviour etc. The framework proposed in this work can be used to recognize such standard movement patterns defined in GIScience literature [1].

1.3 Overview of The Framework

In the world around us, we see movement of many different types of entities. We consider some of these movements as patterns. This may be due to the fact that these movements exhibit some sort of regularity. Otherwise, it may be any arbitrary movement sequence of certain entities that catch our attention and we feel that this movement sequence is important and we should look for its presence. When we think of recognising such movement patterns, then characterisation of a pattern is necessary. This characterisation of a pattern has a subjective element in it. The same sequence of movements of a set of objects can give rise to different notions of patterns. For example, let us consider persons taking a stroll. Let us assume that this movement is somehow interesting to us and we would like to characterize this movement. For someone, it may be enough to note the fact that both of them are moving in a similar direction. Someone else may want additional details and may also like to include the distance between the persons into account. It is still possible that the speed may also come into picture. By saying this, we want to emphasize the fact that a framework for representation of motion patterns should be flexible in characterisation of movement. It should allow addition of new characteristics about a movement pattern whenever necessary. The framework proposed in this thesis has this flexibility. *Elements* or *parameters* of movement have been modelled qualitatively and new qualitative features can be added.

A motion pattern is defined as a formal language in the framework. A motion pattern may have many instances. For example, let us consider movements of two cars where one car overtakes the other. There are many different ways in which this overtake may take place and each of these can be called an instance of the pattern. The overtake may be on right or on left. More patterns may result if we consider the orientation of the overtaken car with respect to the other car after the overtake took place. In a formal language representation, each such instance is represented as a string of the language. This formal language can be learned from training data. Bringing in QSR into the representation has many advantages. QSR makes it possible to use a regular grammar for representation of a motion pattern and accordingly, it is possible to recognise a pattern using finite state automata. One problem with the use of deterministic grammars like regular grammar in learning is that these grammars can not handle the low level errors and missing observations in input data. We have proposed a technique of handling these types of errors using QSR.

Learning a formal language representation of a motion pattern from training data set can be a tedious task. Actually, this is a criticism about grammar based learning techniques. We have designed a qualitative language using which a formal definition of a movement pattern can be provided in the form of a program. Such a program can be interpreted to check for the presence of movement patterns in input data stream. Such a qualitative language has certain other advantages too. Formal grammar based recognition techniques, whether deterministic or probabilistic, have the disadvantage that they can not represent concurrency. In the proposed qualitative language, concurrency within a movement patterns can be represented in the form of temporal constraints between patterns. Another advantage is the ability to construct a pattern hierarchically from already defined movement patterns. Programs written in this qualitative language can be used to define movement patterns for which standard definitions exist. This is particularly true in GIScience where standard taxonomies for motion patterns are available. We have shown how one such taxonomy can be represented in this qualitative language. The framework is also shown to solve a problem of motion pattern recognition of rectangular objects, taking spatial orientation and direction of movement as the movement parameters.

1.4 Contribution

The main contributions of this thesis are as follows:

- The framework presented in this thesis is based on formal algebra. A movement pattern is represented as a formal language and accordingly, recognition is done with the help of finite state automata. Thus, techniques of representation and recognition are not ad hoc and do not vary with application domain. We have been able to confine the representational languages to regular and context free categories. Motion patterns between two objects can be represented as a language that can be parsed using a regular grammar whereas the same among multiple objects can be represented as program elements that can be parsed using context free grammar. An important problem for formal languages is the membership problem. This problem is related with the question "Given a string w and a grammar G, does w belong to the language generated by G?". Use of regular and context free languages is significant because for regular languages, this question can be answered in linear time whereas for context free languages, it can be done in polynomial time. Finite State Automaton (FSA) are recognisers for regular languages. Ability to use finite state automaton makes the task of recognition efficient. Moreover, use of formal languages allows us to represent movement patterns in a hierarchical way.

- Representation of a motion pattern can be done in two ways. Firstly, motion patterns can be learned from a set of training examples. Secondly, definitions of motion patterns can be provided with the help of a program written in a qualitative language designed by us. Both these methods i.e. learning as well as definition using program allow to express concurrency and timing constraints among motion patterns.

- The proposed framework combines QSR with formal languages for representing *elements* or *parameters* of a motion pattern. Use of QSR makes the framework flexible in terms of representation of characteristics of a motion pattern i.e. we can include movement parameters as and when necessary. This makes the framework portable across application domains because in different domains, one may need to use different parameters for characterising a motion pattern. Moreover, it has been shown that it is possible to handle low level errors and missing observations in training data using QSR. Regular and context free languages along with finite automata alone are not sufficient for handling such errors.

1.5 Thesis Outline

In chapter 2, related literature is surveyed. Discussions on QSR and syntactic pattern recognition are presented here. The status of motion pattern analysis research in two key areas, namely, computer vision and GIScience is presented.

Chapter 3 introduces qualitative spatial formalisms for representation of orientation, direction and distance between objects abstracted as rectangles. These formalisms have been used later in an experimental evaluation of the framework for recognition of motion patterns from video samples.

Chapter 4 discusses how QSR can be integrated into syntactic pattern recognition for representation and recognition of motion patterns between two objects. It also explains how grammatical rules can automatically be learned from training data. The issue of handling low level error and missing observations are also addressed.

In chapter 5, design of a qualitative language for representation and recognition of motion patterns is presented. Using such a qualitative language, we can handle multiple objects participating in a pattern. Hierarchical construction of patterns in a bottom up manner becomes possible. Moreover, within a language based setting, concurrency among motion patterns can be represented and recognised.

In Chapter 6, the applicability of the proposed framework in GIScience and computer vision is analysed. Issues of representation and recognition are analysed using real and simulated data.

In chapter 7, we summarize the contributions of the thesis. Limitations of the framework along with possible future extensions are mentioned. At the end of the thesis, the data and the results of various experimental work are presented in four appendices.