Department of Computing

An Investigation and Application of Biology and Bioinformatics for Activity Recognition

Daniel Erwin Riedel

This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University

June 2014

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature: _____ **Date:** _____

Daniel Erwin Riedel

An Investigation and Application of Biology and Bioinformatics for Activity Recognition

by

Daniel Erwin Riedel

Submitted to the Department of Computing
in June, 2014 in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

In recent years researchers have taken particular interest in activity recognition due to the need for automated security surveillance. However, activity recognition can be applied to a plethora of other applications including aged care monitoring, behavioural biometrics, health care monitoring, sports analysis, human computer interaction and smart homes. In a smart home context, sensors can allow monitoring of an individuals health, independence and activities, with the intent to automatically elicit appropriate responses to predetermined or abnormal activities without the need for human supervision, intervention or subjectivity. Such predetermined or abnormal activities could include falling over or remaining seated for lengthy periods of time with possible automated responses including emergency care notification, medical assistance or carer assistance. With an aging worldwide population, development of such smart home technologies can aid the elderly and disabled individuals to maintain their independence, further reducing dependencies on health care systems.

Activity recognition is concerned with the identification and classification of activities, defined as complex events or a series of actions. Automated activity recognition is inherently difficult due to the variable nature with which humans conduct activities. For example, individuals can conduct the same activity with different durations, at different periodic times, in different sequential order, and during other activities (multitasking). Additionally, multi-camera, video-based (spatial) tracking of individuals in a smart home environment is subject to significant variability particularly in multi-room residences. To address the issue of spatial noise from video tracking systems and to detect activities occurring concurrently, this thesis proposes a novel biologically-inspired cellular chemotactic model. The cellular chemotactic model, which is based on bacterial chemotaxis, a mobility survival characteristic employed by certain bacteria in diverse

environments, represents activities as chemotactic cells in a two dimensional space. As activities are matched the cells conduct a biased random walk of a longer duration towards an attractant source (improving cell "fitness"). In the presence of noise, cells conduct shorter unbiased random walks (robustness mechanism). The activity type for a target sequence is determined by finding the activity cell with the smallest Euclidean distance to the attractant origin. Since multiple cells can move independently of each other (similar to agent-based models), the approach allows recognition of concurrent and interwoven activities (multi-tasking).

The recognition of spatial activities with tracking noise and temporal variation is non-trivial. Discretisation of two-dimensional spatial data to form symbolic representations for activity recognition can be problematic as discretisation can affect the model parameters and accuracy. Two-dimensional approaches are more accurate however typically suffer from susceptibility to noise. This thesis analyses sequence alignment approaches and proposes two robust sequence alignment approaches (Longest Common Subsequence Distance (LCSD) and Global Edit Distance (GED)) for spatial activity recognition, and a robust and more computationally efficient means for recognition of spatial sequences in the presence of temporal variation (Threshold Dynamic Time Warping (TDTW)). The LCSD and GED approaches are demonstrated in this thesis to be robust to innate sequence variability and noise from spatial sequences. The TDTW approach is also shown to be capable of accurately recognising sequences with temporal variation and in the presence of noise, outperforming popular template and probabilistic approaches.

Online video tracking systems provide a continuous stream of spatial sequences, which are typically processed using a sliding window approach. Existing activity recognition techniques segment these streams at likely activity "signatures" for computational efficiency, and analyse the respective sequence windows for embedded activities. Many of these approaches determine sequence similarity based on the whole window sequence and are unable to detect embedded sub-sequences. This thesis proposes an algorithm based on Smith-Waterman (SW) local alignment from the field of bioinformatics that can efficiently locate and accurately quantify embedded activities within a windowed sequence. The SW local alignment technique is also robust to missing sequence information and tracking system noise. Another variant of the approach called Online SW (OSW) allows for efficient calculation of the SW local alignment with continuous spatial streams. Experiments further demonstrate that the technique successfully recognises embedded activities with a high degree of discimination.

Acknowledgments

This dissertation would not have been possible without the following people:

My supervisors, Prof Svetha Venkatesh and A/Prof Wanquan Liu, for their enduring support, expert guidance and understanding throughout the course of this research.

My friends and colleagues at the Department of Computing in particular Patrick, Simon, Thomas, and Thorsten who provided advice, friendship and humour throughout this journey.

My family who provided inspiration and support over the lengthy duration.

Published Work

This dissertation is based upon several works that have been published over the course of the author's PhD. These referred journal and conference publications are as follows:

Riedel D.E., Venkatesh S, and Liu W, 2008, Recognising spatial activities using a bioinformatics inspired sequence alignment approach. *Pattern Recognition*, vol. 41, no. 11, pp. 3481-3492.

Riedel D.E., Venkatesh S, and Liu W, 2007, Threshold dynamic time warping for spatial activity recognition. *International Journal of Information and Systems Sciences*, vol. 3, no. 3, pp. 392-405.

Riedel D.E., Venkatesh S, and Liu W, 2006, A chemotactic-based model for spatial activity recognition. *International Journal of Systems Science*, vol. 37, no. 3, pp. 949-959.

Riedel D.E., Venkatesh S, and Liu W, 2006, A Smith-Waterman local sequence alignment approach to spatial activity recognition, In *Proceedings of the IEEE International Conference on Video and Signal Based Surveillance (AVSS06)*, IEEE, Sydney, Australia, pp. 54-59.

Riedel, D.E., Venkatesh S, and Liu W, 2005, Spatial activity recognition in a smart home environment using a chemotactic model. In *Proceedings of the IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP05)*, Melbourne, Australia, pp. 301-306.

Contents

Acknowledgments **i**

Published Work **ii**

Contents **v**

Figures **viii**

Tables **ix**

1 Introduction **1**
- 1.1 Aims and Approach . . . 3
- 1.2 Significance and Novelty . . . 5
  - 1.2.1 A biological paradigm for spatial activity recognition . . . 5
  - 1.2.2 Improving robustness of spatial activity recognition using bioinformatics principles . . . 6
  - 1.2.3 Recognition of activity sequences in the presence of noise and temporal variation . . . 7
  - 1.2.4 A bioinformatics approach for detecting embedded spatial sequences 8
- 1.3 Structure of the Thesis . . . 9

2 Background and Related Work **11**
- 2.1 Human Activities and Characteristics . . . 11
- 2.2 Smart Homes . . . 14
- 2.3 Human Activity Recognition Approaches . . . 16
  - 2.3.1 Template Matching Techniques . . . 18
    - 2.3.1.1 Dynamic Time Warping (DTW) . . . 19
    - 2.3.1.2 Edit Distance on Real Sequence (EDR) . . . 21
    - 2.3.1.3 Edit Distance with Real Penalty (ERP) . . . 22
  - 2.3.2 Hidden Markov Models . . . 23
- 2.4 Bioinformatics: Sequence Alignment . . . 24
  - 2.4.1 Longest Common Subsequence (LCSS) . . . 29

2.4.2 Edit Distance . . . 30
2.4.3 Smith-Waterman Local Alignment . . . 31
2.5 Biologically-Inspired Computational Models . . . 33
2.6 Activity Data Sets . . . 34
2.6.1 Camera Tracking System . . . 34
2.6.2 Dataset A (10 activities) . . . 38
2.6.3 Dataset B (3 Activities) . . . 42
2.6.4 Dataset C (12 activities) . . . 42
2.6.5 Discretisation of Spatial Sequences . . . 43
2.7 Summary . . . 44

3 A Chemotactic Paradigm for Recognising Simple and Interwoven Spatial Activities 45
3.1 Bacterial Chemotaxis . . . 46
3.2 Cellular Chemotactic Model . . . 48
3.3 Methodology . . . 54
3.4 Experimental Results . . . 56
3.4.1 Parameter Selection . . . 56
3.4.2 Recognition Performance . . . 58
3.4.3 The Effect of Training Size on Recognition Performance . . . 58
3.4.4 Noise Tolerance . . . 60
3.4.5 Recognising Simple Interwoven Activities . . . 60
3.5 Summary . . . 64

4 Improving robustness of spatial activity recognition using principles from bioinformatics 65
4.1 Nomenclature . . . 66
4.2 Longest Common Subsequence Distance (LCSD) . . . 66
4.3 Global Edit Distance (GED) . . . 70
4.4 Evaluation of LCSD and GED . . . 71
4.4.1 Parameter Selection . . . 74
4.4.2 Recognition Performance of LCSD and GED . . . 75
4.4.3 Training Set Size versus Recognition Performance . . . 80
4.4.4 Effect of Noise on Recognition Performance . . . 80
4.5 Summary . . . 82

5 Recognising activity sequences in the presence of noise and temporal variation **84**
5.1 Threshold Dynamic Time Warping (TDTW) 85
5.1.1 Band DP constraint 88
5.2 Evaluation of Threshold Dynamic Time Warping (TDTW) 89
5.2.1 Effect of Noise on Recognition Performance 91
5.2.2 Effect of the Band-DP window size on Recognition Performance . 96
5.3 Summary 98

6 Recognising Embedded Activities within Spatial Sequences **100**
6.1 The Smith-Waterman (SW) Approach 103
6.2 Online Smith-Waterman (OSW) 107
6.3 Experimental Results 114
6.3.1 Parameter Selection 115
6.3.2 OSW, DTW and HMM with Online and Inaccurate Activity Segmentation 117
6.3.3 SW, DTW and HMM with Accurate Activity Segmentation 119
6.3.4 Robustness of SW, DTW and the HMM with Accurately Segmented Activities 122
6.4 Summary 125

7 Conclusion **126**
7.1 Future Directions 129

Bibliography **141**

List of Figures

2.1 Euclidean distance with temporal warping. 18
2.2 DTW with temporal warping. 19
2.3 An example alignment of two short sequences. The symbol - refers to an *indel* and represents an insertion or deletion in either sequence. Two gaps are present in the alignment; one of size 1 and the second of size 2. 25
2.4 An example sequence alignment containing gaps of size one and two. . . . 27
2.5 Schematic global and local alignments between two sequences. 27
2.6 Layout of the Smart Home environment. 35
2.7 Room 1 of the Mock Smart House Environment. 35
2.8 Configuration of the distributed camera tracking system (from Nguyen *et al.* (2002)) 36
2.9 An example of 2 segmented blobs post-background subtraction and blob merging 37
2.10 CPM and CM processing with the distributed camera tracking system (from Nguyen *et al.* (2002)) 38
2.11 Spatial paths of dataset A activities. 41
2.12 Spatial paths for dataset B activities. 42
2.13 Mapping of smart house trajectories to symbolic forms. 43
2.14 Discretisation Grid for Room 1 of the Mock Smart House Environment. . 44

3.1 Cellular Chemotactic Environment E outlining the origin for cells c, the attractant origin s and the high concentration region governed by θ_{high} . . 49
3.2 Graphical representation of the symbol matching process at $t = 0$ (initialisation) 50
3.3 Graphical representation of the symbol matching process at $t = 1$ (match) 51
3.4 Graphical representation of the symbol matching process at $t = 3$ (non-match) 51
3.5 Example of a short random walk behaviour when cells c are close to an attractant ($d \leq \theta_{high}$) 53
3.6 Example of a return to normal behaviour when cells c move away from an attractant ($d > \theta_{high}$) 53

3.7 Chemotactic Cell Movements of true and false class cells. Movement is in the direction of the origin. 54
3.8 Number of training sequences versus accuracy for the HMM and chemotactic models. The bars represent one standard deviation from the mean. 59
3.9 Noise magnitude versus classification accuracy for the HMM and chemotactic models. The bars represent one standard deviation from the mean. 61
3.10 Snack-Watch TV and Cooking-Eating spatial paths. 62
3.11 Interwoven cell movements (a) Snack-Watch TV Cell (SWTV) (b) Cooking Cell. 63

4.1 Activities exhibiting temporal variation (x axis represents the sequence length). 73
4.2 GED γ Parameter Optimisation. 75
4.3 Recognition performance for LCSD, GED compared to LCSS, DTW and the HMM. 76
4.4 Conceptual representation of how DP-based techniques form non-overlapping warping alignment paths in contrast to the HMM which creates a single generalised alignment path. 79
4.5 Training set size versus accuracy (%) for datasets A and B. 81
4.6 Noise magnitude versus accuracy for dataset B. 82

5.1 Windowing effect of θ on a two dimensional x, y coordinate space. 86
5.2 An optimal warping path for **a** and **b**. 87
5.3 Application of a Band DP constraint with window size m. 89
5.4 Recognition performance vs number of training sequences for the discrete HMM ($M = 156, N = 5$) 91
5.5 Average difference in DTW and TDTW intraclass distances (same activity) as a result of noise. 92
5.6 DTW and TDTW Class distances. 94
5.7 Band DP width m versus precision and recall. 97
5.8 Band DP width m versus relative runtime. 98

6.1 Online Spatial Activity Recognition using a Sliding Window. 101
6.2 The affect of θ on spatial activity recognition. 105
6.3 Example sequences **a** and **b** 106
6.4 An optimal SW alignment using sequences **a** and **b**. 107
6.5 Initialisation of the DP matrices using OSW. 108

6.6 OSW recognition with a window size of w=4. 113
6.7 Empirical SW Parameter Optimisation. 116
6.8 HMM number of hidden states N versus classification accuracy. 117
6.9 Confusion matrix for SW. The legend represents the percentage of activity sequences classified. 120
6.10 Spatial patterns for confused activities 4 and 5. 121
6.11 Noise magnitude versus classification accuracy. 124

List of Tables

2.1 Characteristics of datasets A,B and C used in this thesis. 39
2.2 Dataset A activities and actions. 40

3.1 Classification accuracy of discrete HMM models with $M = 5, 7, 12$ and 15 hidden states. 58
3.2 Classification accuracy for the Chemotactic and HMM models. 58

5.1 TDTW C matrix using example sequences **a** and **b** with $\theta = 1.0$m. 87
5.2 DTW C matrix using example sequences **a** and **b**. 87
5.3 Precision and recall rates for threshold-based NN classification using DTW, TDTW and the discrete HMM. 95

6.1 SW C matrix using example sequences **a** nd **b**. 106
6.2 Incremental Derivation of DP Matrices 109
6.3 Threshold-based NN classification with Online Recognition. 118
6.4 Threshold-based NN classification with Accurate Activity Segmentation. . 120

Chapter 1

Introduction

Activity recognition is concerned with the identification and classification of complex events or series of actions. It is a multi-faceted problem domain primarily including aspects of computer vision and pattern recognition. Activity recognition can be applied to areas including security surveillance, aged care monitoring, behavioural biometrics, health care monitoring, sports analysis, human computer interaction and smart homes (Dilger, 1997; Aggarwal and Cai, 1999; Ehlert, 2003; Aggarwal and Ryoo, 2011). Each activity recognition application has unique environmental conditions and constraints, requiring novel solutions or optimisation of existing solutions for that application. For example, video-based activity recognition in a metropolitan crowd monitoring scenario typically utilises a macro-view of a scene to determine user activities and behaviour (Andrade *et al.*, 2006; Fuentes and Velastin, 2006; Robertson and Reid, 2006) . These approaches need to deal with smaller amounts of information to categorise an individuals activities, with a high degree of scene and crowd variability especially with outdoors environments. In contrast, a video-based HCI solution in an office environment focuses on a micro-view of a user (limbs, extremities and body characteristics), rather than a scene, to categorise user activities based on body movement (Ben-Arie *et al.*, 2002). The challenge for researchers in this scenario is accounting for an individuals variability in motion and characteristics for each activity and between different activities.

In the context of security surveillance, activity recognition systems could detect individuals acting suspiciously and consequently alert security forces without the need for constant human monitoring of surveillance cameras. With sports analysis, activity recognition could be used to automatically capture statistics on an individual's performance. In a smart home, sensors could allow monitoring of an individual's health, independence and activities, with the intent to automatically elicit appropriate responses to predetermined or abnormal activities. A study by Hine *et al.* (2005) shows a direct correlation between daily living activities conducted by a person and their well-being.

Thus, changes in a person's everyday activities is indicative of a change in well-being or health. In an aged care scenario activity recognition systems could be used to identify person's remaining seated for lengthy periods. Such a system could then provide an automated response such as a medical or emergency services notification or request for carer assistance. With an aging worldwide population, development of such smart home technologies can assist patients, the elderly and disabled individuals to maintain their independence, further reducing dependencies on already overloaded health care systems.

Activity recognition is complex due to the variable nature with with humans conduct activities. For example, individuals can conduct the same activity over different durations, at different periodic times, in different sequential order, and during other activities (multi-tasking). Under these conditions, humans can readily generalise and classify these activities as the same type; this equivalent process has been difficult to replicate in artificial intelligence and pattern recognition approaches. If one makes the assumption that a sensor system can accurately capture positional, pose, and other information when an activity occurs, activity recognition approaches still need to generalise or classify all the activity variants into a singular activity class, which is not trivial (Szalai, 1972; Tapia, 2003; Sheikh *et al.*, 2005). Realistically sensor systems such as multi-camera video tracking systems produce data with significant noise, particularly in multi-room residences. This noise makes correct classification of activities that exhibit temporal or spatial intra-class variability more difficult.

Activity recognition researchers have developed numerous template matching, probabilistic, stochastic and hybrid models to address the issues of recognising human activities. These models are typically constrained to niche applications and environments, and can suffer from problems including low recognition performance with the variable nature of activities, high computational complexity and/or poor robustness in real world applications. The reviews by Shah and Jain (1997); Aggarwal and Cai (1999); Tapia (2003); Turaga *et al.* (2008) and Aggarwal and Ryoo (2011) provide a thorough coverage of activity recognition approaches with several different taxonomies provided.

This thesis accepts the premise that activity sensor information is inherently noisy and variable, and explores alternative pattern recognition paradigms that are tolerant to spatial sensor and activity variability, yet discriminate. The spatial sensor domain is a focus of this research due to the ease and non-invasive manner with which sensor data can be collected in a smart home environment. For each activity, a sequence

of (x,y) tuples are produced based on a person's relative position in the smart home environment. These sequences are in turn labelled with the activity type in a supervised manner. This approach of representing activities as spatial sequences is consistent with activity recognition approaches in other smart home studies (Lühr *et al.*, 2003; Nguyen *et al.*, 2003; Duong *et al.*, 2005; Zouba *et al.*, 2007).

1.1 Aims and Approach

Video-based activity recognition is challenging due to tracking difficulties with dynamic environments, the large state space associated with different human activity types and the variable nature of human activities. The activity recognition problem space in this thesis is constrained to a smart homes environment, limiting tracking to a consistent indoor environment, with a small number of individuals. The constrained indoor environment also limits the number of activities that can be conducted by individuals. The underlying assumption that sensors can capture humans conducting activities in a consistent and period manner has been empirically verified in studies by Monk *et al.* (1992); Suzuki *et al.* (2004) and Suzuki *et al.* (2006). In this thesis, non-invasive camera sensors are utilised, which are processed via a multi-camera tracking system Nguyen *et al.* (2003) to provide relative spatial coordinates of individuals in the smart home environment.

The above constraints make the activity recognition problem space tractable for analysis and development of new approaches for automatic recognition. The following outlines the key aims of this thesis with respective approaches given below:

Aim 1. To identify, validate and apply new paradigms to address shortfalls of existing spatial activity recognition approaches in a smart homes context.

Aim 2. To develop robust spatial activity approaches that can recognise activities in the presence of human variation and noise generated by video-based tracking systems in a smart homes context.

Aim 3. To develop discriminate and efficient spatial activity approaches, that can recognise embedded spatial activities in a continuous stream of spatial data, generated by a video-based tracking system.

In this thesis and to address the first aim, a biological process is identified and modelled to generate the cellular chemotactic model for spatial activity recognition. This model incorporates aspects of bacterial chemotaxis, a robust biological process that allows bacterial cells to directionally swim in response to chemical or other physical gradients, thus improving survivability in dynamic environments. The application of the cellular chemotactic model to the spatial activity domain demonstrates similar robustness characteristics to its biological process, allowing recognition of spatial sequences in the presence of noise and concurrent recognition of multiple simple interwoven activities (a multi-cellular chemotactic trait).

Bioinformatics is a useful source of inspiration for dealing with the variable nature of human activities and addressing the second aim. The sequence alignment techniques used with biological sequences addresses similar issues to those found with human activities. These sequence alignment issues (and their corresponding activity related issues) include sequence compression (relating to shorter duration activities), sequence expansion (relating to longer duration activities), sequence insertions and deletions (relating to activity variability and tracking noise) and sub-sequence matching (relating to activity interweaving) (Sankoff and Kruskal, 1999a). In this thesis, sequence alignment is adapted to the two dimensional space and evaluated with existing activity recognition approaches. The result of this analysis is the fusion of characteristics of sequence alignment methodologies with "time warping" philosophies. The longest common subsequence distance (LCSD) and global edit distance (GED) are in turn formulated for dealing specifically with noise tolerance and minor temporal activity variation. A threshold DTW (TDTW) approach is also specifically developed for recogition of activities exhibiting temporal variation, whilst having an improved tolerance to tracking system noise.

The issue of recognising embedded spatial activities and addressing the third aim is approached through the use of another bioinformatics inspired approach. The Smith-Waterman (SW) local alignment technique is typically used in bioinformatics to search for genes and fragments (subsequences) in unknown sequences. Application of the SW technique to spatial activity recognition provides a unique ability to identify and quantify sub-sequences (corresponding to activities) in spatial data streams. A two dimensional SW approach was consequently developed for the spatial domain, based on the original SW local alignment algorithm, and evaluated against existing sequence alignment and activity recognition approaches. Improved computational efficiency was achieved with the SW algorithm, through investigation of sub-optimal characteristics in dynamic pro-

gramming with SW. This resulted in the formulation of the online SW (OSW) approach to spatial activity recognition.

1.2 Significance and Novelty

This thesis makes four main contributions to the field of pattern recognition:

1. The development of a novel biological paradigm (chemotaxis) for robust and concurrent pattern recognition in an activity recognition context.
2. Adoption of bioinformatics sequence alignment characteristics to improve recognition of spatial activities in the presence of noise.
3. Fusion of time-warping and bioinformatics sequence alignment characteristics to improve robustness and allow recognition of the same activities with temporal variation.
4. The novel use and optimisation of Smith-Waterman local alignment for recognition of embedded spatial activities in a spatial data stream.

The thesis contributions and their significance are as follows.

1.2.1 A biological paradigm for spatial activity recognition

Activities captured by multi-camera tracking systems and represented as spatial sequences in a smart home environment are traditionally noisy, exhibiting significant variation due to tracking artifacts (shadows, reflections, lighting variation, occlusions). As a smart home and its various rooms are small and the tracking noise can be large (even with smoothing), these variations can affect activity discrimination.

Modelling this spatial activity recognition problem using a biological paradigm is novel. Biological systems have proven to be a useful basis for solving many real world problems as a consequence of their innate robustness, adaptiveness, diversity and error tolerance

(Paton, 1994). The proposed chemotactic approach for spatial activity recognition is significant as it can address the robustness issue with noisy spatial activity sequences. Bacteria that exist in competitive and nutrient poor environments, are believed to have evolved the chemotactic capability to sense and respond to dynamic environments, increasing their survivability and thus evolutionary fitness. Through the use of a chemotactic paradigm, a cellular chemotactic model has been developed that is capable of dealing with significant noise from video tracking systems. The tolerance to tracking noise is achieved in the cellular chemotactic model through the process of biasing random walks (over longer durations) of activity cells towards an "attractant" where subsequences correspond, whilst conducting unbiased (shorter duration) random walks for subsequences that don't match. At the conclusion of a test sequence comparision, the cell that clost to an "attractant" is determined to the be activity class. As the cellular chemotactic model is not restricted by Markovian constraints, unlike HMM-based techniques, the chemotactic approach also exhibits inherent resilience to spatial variations in activity sequences of similar duration.

One of the most difficult aspects of activity recognition is the ability to recognise interwoven or multi-tasked activities. This is difficult to model from a computational perspective as humans can interweave activities at any time and between most activities, making *a priori* information on interwoven activities in a supervised or semi-supervised approach of low value. The cellular chemotactic model is capable of addressing this multi-tasking recognition issue as to its cells mimic the behaviour of agent-based models, with multiple activity cells responding to a spatial activity sequence. This characteristic allows the cellular chemotactic model to cater for interweaving of simple activities.

1.2.2 Improving robustness of spatial activity recognition using bioinformatics principles

Spatial sequence variation as a result of tracking noise can affect activity discrimination and recognition classification performance. Tracking noise typically manifests as a localised variability, compression or expansion of sequences. In biological sequences, localised substitution, compression and expansion of sequence elements are naturally occurring phenomena in nature. As a result bioinformatics techniques have been optimised to identify known gene and protein sequences exhibiting variation. The proposed

LCSD and GED approaches addressed in this thesis are based on bioinformatics sequence alignment techniques that are capable of quantifying sequence variability. The LCSD approach uses real spatial data to locate optimal sequence alignments and assigns a distance score between a template sequence and a test sequence. This is achieved by minimising the dissimilarity between sequence elements over the length of the sequences. The GED approach differs in that it finds the optimal alignment and distance between a pair of sequences by minimising the number of mismatches, insertions and deletions with differing penalties over the sequence length. The LCSD and GED approaches are explored in the context of spatial activity recognition through evaluation with spatial sequences containing artificially introduced noise, and contrasted to existing template-based and probabilistic spatial activity recognition approaches. This contribution is significant as the application of bioinformatics inspired methodologies to the spatial activity recognition domain is shown to provide robust activity recognition.

1.2.3 Recognition of activity sequences in the presence of noise and temporal variation

In order for an automatic activity recognition system to accurately recognise human activities, the system must be able to cater for temporal variations in the same activity. Humans characteristically conduct the same activity over differing durations. For example, the activity of having breakfast in a smart home can occur over two distinct time periods depending on whether the individual is in a rush to carry out another activity or not. Current activity recognition approaches are capable of recognising activities with temporal variation, yet are highly susceptible to noise from tracking systems.

Recognition of spatial activities exhibiting temporal variation in the presence of noise, is addressed in particular by the threshold DTW (TDTW) approach. TDTW is significant as it tolerant to inherent noise resulting from video-based human tracking, yet is capable of matching the same activities with temporal variations. Noise tolerance of this approach is achieved through the introduction of a novel sequence alignment distant matching constraint in the TDWD calculation for spatial elements. The constraint prevents minor warping with small changes in position, reducing the algorithms susceptibility to tracking noise. The improved discrimination ability of TDTW in the presence of noise is verified against existing sequence alignment and activity recognition approaches

across multiple data sets, demonstrating the superior performance of the approach. To improve the runtime performance of TDTW a band dynamic programming (DP) constraint is also introduced and validated in the context of spatial activity recognition of segmented activities. The band DP constraint does reduce algorithm runtime with only minor decreases in recognition performance with smaller band sizes.

1.2.4 A bioinformatics approach for detecting embedded spatial sequences

An automatic activity recognition system must be able to recognise human activities in continuous streams of data. In traditional activity recognition approaches, data streams are processed using a sliding window approach where the window size corresponds to an activity duration or "signature". The extracted and segmented sequence is then analysed to determine sequence similarity or applied to a model for recognition. Sliding windows of different sizes need to be provisioned for recognising different activities if they have different lengths as most activity recognition approaches are unable to detect activities embedded in a larger sequence.

The Smith-Waterman (SW) local alignment algorithm Smith and Waterman (1981) is a bioinformatics dynamic programming approach that compares two biological sequences, finds the optimal sub-sequence in relation to a penalty function, and provides a similarity metric of two sub-sequences. In this research, the SW algorithm is adapted to a two dimensional space, a linear gap penalty based on Euclidean distance is used between points rather than a constant, and a matching constraint is applied for the spatial activity domain. The use of a bioinformatics inspired approach in the spatial activity recognition domain is novel. The two dimensional SW approach allows one to efficiently locate and quantify similarity of embedded spatial activity sequences in a spatial data stream, as well as detect optimal sub-sequences with only partial activities. This ability to identify matching sub-sequences results from the algorithm penalising and terminating poorly aligned or mis-matched sequences. The ability to recognise spatial activity sequences from online video tracking systems is significant as traditional approaches focus on applying full window sequences to models or template recognition. Alternatively, existing activity recognition algorithms are heavily reliant on accurate stream segmentation for recognition of spatial activity sequences. The two-dimensional SW ap-

proach provided in this thesis does not suffer from these limitations, and furthermore has been optimised (see online SW (OSW)) for continuous processing of spatial streams from video tracking systems. Experimental validation with existing sequence alignment and activity recognition approaches confirms the high recognition performance of the developed approach.

1.3 Structure of the Thesis

This thesis is organised as follows:

Chapter 2 describes human activities and their characteristics in a smart home context, followed by a review of the related work in the fields of human activity recognition, sequence alignment, and biologically-inspired computational models. Existing methods of activity recognition are briefly explored with detailed analysis provided for similar approaches or approaches used in validation. A brief review of biologically inspired models is provided to allow discussion in other relevant chapters and future work. A brief synopsis of the smart home laboratory environment, the tracking system utilised, and the experimental data sets is also included.

Chapter 3 outlines the biological process of chemotaxis, and how the process is abstracted to form a cellular chemotactic activity recognition model. The cellular chemotactic model is then outlined, followed by the experimental methodology for evaluation, empirical parameter optimisation and experimental validation against a discrete HMM in regards to noise tolerance and recognising interwoven activities.

Chapter 4 describes LSCD and GED sequence alignment approaches that are explored in the context of robust spatial activity recognition in the presence of tracking system noise. The experimental methodology for evaluation and empirical parameter optimisation of LCSD and GED approaches are given. These approaches are also validated against other sequence alignment and probabilistic approaches.

Chapter 5 describes a TDTW approach which combines bioinformatics principles with time-warping methodologies to improve noise tolerance with spatial sequences exhibiting temporal variation. The experimental methodology for evaluation, empirical parameter

optimisation of TDTW is provided and the approach is validated against other sequence alignment, probabilistic and time warping algorithms.

Chapter 6 outlines a novel Smith-Waterman (SW) local alignment approach to recognition of embedded spatial sequences in an online video recognition system. A more efficient online variant of the two dimensional Smith-Waterman algorithm, termed Online SW (OSW), is also described and evaluated. Factors affecting optimal SW and OSW parameter estimation are covered, in addition to experimental validation against DTW and the HMM. Further experiments are conducted to analyse the discriminatory performance of the two dimensional SW approach in comparison to DTW and the HMM for accurately segmented spatial sequences.

Chapter 7 provides a summary of the research in this thesis and potential avenues for future research in spatial activity recognition and smart homes.

Chapter 2

Background and Related Work

The investigation of activity recognition that is undertaken in this thesis is concerned with recognising spatial activity patterns in the context of smart homes. Much of its inspiration is drawn from biological paradigms and bioinformatics techniques to formulate novel pattern recognition approaches that are described in Chapters 3 - 6. This chapter reviews the related literature for the approaches developed in this thesis, and familiarises the reader to the activity data sets collected and used in evaluations.

The chapter is organised as follows: Section 2.1 discusses human activities and aspects of their innate variability, followed by a review of Smart Homes in the context of activity recognition in 2.2. In Section 2.3 a review of related work in the field of human activity recognition is provided, which also includes a brief outline on approaches evaluated in this thesis. The section focuses on spatial activity recognition approaches, which are otherwise called trajectory-based activity recognition or movement recognition in the literature. A review of bioinformatics and sequence alignment approaches, which are inspiration for some of the research in this thesis, are provided in 2.4. In Section 2.5 a discussion is provided on biologically inspired models and paradigms, which motivated formulation of a chemotactic-inspired model. An outline is then given in Section 2.6 of the smart home laboratory environment, the tracking system utilised, and the experimental data sets that were collected and evaluated. The chapter concludes with a summary in Section 2.7.

2.1 Human Activities and Characteristics

A human activity is defined as a series of actions taken in pursuit of a objective, whereby a objective may be to seek nourishment or sleep. Human activities are influenced by a

myriad of factors and can be carried out in an infinite number of ways. The fundamental assumption underlying human activity recognition is that humans conduct activities in a consistent manner, albeit with some variability. This assumption is required in order to generalise or categorise known activities into labeled classes, to then allow classification of observed activities in relation to known classes. Much of the activity recognition literature makes the assumption that human activities are consistent without validating the assumption through empirical studies, or understanding the factors that may affect the variability of activities. Understanding the causes of activity variability allows one to estimate the expected level and type of variability of normal activities, which can assist with optimisation of activity recognition approaches.

In the context of activities of daily living (ADL) which are the everyday activities that one would carry out, Hine *et al.* (2005) formulates a conceptual framework of factors that affect well-being and in part ADLs. Well-being in this instance being a combination of human mental, social and and physical states that determine an individual's quality of life. In the framework, the authors predicate that contextual and personal factors, combined with an individual's hobbies or interests affect ADLs and hence well-being. Contextual factors include ones home environment, social network, locale and social support, whilst personal factors include physical and psychological attributes. Changes in any of the contextual factors, personal factors or hobbies / interests, can affect ADLs including the manner and duration over which they are conducted. The intrinsic variability of these factors for individuals with normal well-being is yet to be empirically determined. Quantifying these influences on human activities is important to estimate the likely variability to be expected for normal activities for activity modelling.

Studies of individuals in technology enabled smart homes or facilities has enabled empirical analysis of human activity patterns. In the study by Suzuki *et al.* (2001, 2004, 2006), ADL patterns of non-elderly and elderly people are captured over days to months to determine the correlation of recorded sensor patterns to patient recorded ADL's and thus wellbeing. Studies were conducted using combinations of non-invasive sensors including infrared, door, window, photoelectric, thermal, gas usage and wattage. These empirical studies demonstrate that sensor patterns do strongly correlate to patient recorded ADL's and well-being in real world scenarios. Importantly, the results also show that different individuals exhibit different ADL routines and activity durations; however, ADL routines and activity duration are consistent for an individual at a normal level of well-being. In Suzuki *et al.* (2004), the authors were also able to infer a subject's well-being from

changes to their normal ADL routine, which is important for health care monitoring.

The landmark study of daily activities by Szalai (1972) at a multi-national and population perspective showed that human activities are affected by location, cultural habits, time and the order of activities. Location determines the range of tasks that can be conducted and can be seen as a constraint on the search space of activities. Location also influences the activity actions. For example, *working* at home will have a different set of actions to *working* in the office. Cultural habits can influence the timing, duration, and type of activities. Using the example of working, some cultures have a work day that commences at 8am and finishes at 6pm, whilst others start at 9am and finish at 5pm. Cultural habits may also mandate that certain activities be conducted at particular times of the day, for instance the activities of having morning and afternoon tea. Time has numerous effects on the conduct of activities. In Szalai (1972), periodic variations relating to the day, day in the week (weekday versus weekend), week, month, and season determine which activities are conducted, and activity duration. For example, during weekdays the activity of breakfast is typically shorter than breakfast on a weekend, and in summer the activity of exercise may be for one to go for a run outdoors, but in winter the activity of exercise maybe to go to the gym. Time also has an impact on multi-tasking of activities, where activities are conducted concurrently or interwoven to obtain overall efficiencies in time or resources. Multi-tasking of activities is also prevalent in situations where an individual has a large number of objectives and a short duration with which to finish them. Lastly, relationships between activities may determine or restrict the order that activities may be conducted. For example, the activity of *working* must follow the activity of *driving to work*, and be followed by *drive homing from work*.

By combining the framework of Hine *et al.* (2005) with the findings of Suzuki *et al.* (2001, 2004, 2006); Szalai (1972), key influences of human activity variation can be ascertained. These influences are as follows:

- Contextual Factors
  - Home environment*
  - Locale or location*
  - Social Network
  - Social Support

  - Cultural habits*
  - Time and periodicity*
  - Order of activities*
- Personal Factors
  - Physical attributes
  - Psychological attributes
- Hobbies and interests

The contextual factors highlighted with a * are those that can be captured non-invasively with sensors or can be provided through *a priori* information on locations, times or cultures. The personal factors and information on an individual's hobbies and interests require more invasive means. The degree of influence each of these factors have on human activity variation has not been quantified. Without understanding these influences and their affect on human activities, it is very difficult to design activity recognition approaches that could accurately recognise human activities through the innate variability. The localisation of human activities to a smart home environment does reduce the complexity of this task and non-invasive sensor approaches can be used to recognise human activity patterns as shown by Suzuki *et al.* (2001, 2004, 2006). Unfortunately, the robustness of activity recognition approaches will always be limited due to the complexity of factors affecting human activities.

2.2 Smart Homes

A smart home is a residence that is fitted with sensor and/or assistive technologies to provide automated monitoring and/or assistance to its occupants. An aging global population, with increasing health care requirements is the key driver for research in this field. It is envisaged that smart homes will automatically monitor elderly and patient lifestyle, allowing individuals to maintain a higher quality of life in an independent capacity and to provide notification to health care support services when required. Sensors used in smart homes can be either non-invasive (video, audio, contact, temperature, switch, and infrared) or invasive (pedometers, accelerometers, GPS, RFID, mobile devices, implantable devices and microcapsule devices). Non-invasive sensor approaches

are more popular with inhabitants as they are less intrusive on lifestyle and individuals do not have to remember to place the sensor on their body prior to going about their daily business (video and audio sensors are still seen by some to be invasive in regards to privacy). Invasive sensors were originally large, limited in battery life and intrusive as seen in Lee and Mase (2002); however, with the advent of smaller wearable technologies such as smart watches becoming more ubiquitous in society, there has been an increased interest in this field of pervasive computing. The following elaborates on the Welfare Techno-House (WTH) smart home as covered in Suzuki *et al.* (2001), CareMedia (Hauptmann *et al.*, 2004), the intelligent Dormitory (iDORM) (Hagras *et al.*, 2002; Rivera-Illingworth *et al.*, 2005) and the CASAS smart home described in Fang *et al.* (2012). The reader is directed to Chan *et al.* (2009) for a thorough review of smart home literature and relevant assistive technologies.

The WTH Mizusawa smart home (Suzuki *et al.*, 2001) is a one bedroom apartment fitted with infrared, door, window, and appliance sensors that record a binary value at the time of an event. The sensors are placed on all taps, doors, windows, and appliances, with infrared sensors installed in each room to detect the presence of a singular individual and interactions. WTH Mizusawa is constrained to only being able to monitor a singular inhabitant due to the granularity of its sensor data and its inability to differentiate between inhabitants. The binary nature of the sensor data also limits the ability of the smart home to be used for discrimination of more complex activities, but is ideal for recognising simple activities such as sleeping or having a meal.

The CareMedia project (Hauptmann *et al.*, 2004) is employed in a nursing home using non-invasive video sensors (1 per room) to monitor inhabitant movements. The inhabitants are tracked through multiple rooms in the nursing home using video-based background segmentation (due to the stable lighting environment and fixed cameras) with region growing and noise removal to minimise tracking artifacts. Individuals are correlated to regions according to their colour histograms, with RANSAC used to determine motion patterns of body parts when individuals are static. Unlike the WTH smart home, CareMedia allows simultaneous tracking of multiple inhabitants, but lacks the fidelity to determine human interactions resulting from inclusion of appliance sensors.

iDorm is a one-bedroom dormitory style smart home fitted with various sensors including temperature, system monitors, light and pressure, as well as effectors including door actuators and equipment switches (Hagras *et al.*, 2002; Rivera-Illingworth *et al.*,

2005). Unlike the other smart homes and systems described here, the iDorm provides the inhabitant with a user interface with which they can inquire about the state of the environment and modify the system state in response to activities (to assist with automated learning). The iDorm is only able to accurately monitor a single inhabitants activities due to the granularity of the sensors employed, similar to the WTH smart home.

The CASAS smart home (Fang *et al.*, 2012) is a three bedroom apartment fitted with a 1-metre grid of motion sensors to monitor inhabitant movements within the home, correlating these to known activities. Post collection of data, sensor events are annotated for ADLs which are used to train the generative models for ADLs. The processed sensor information is represented by the sensor ID, time of day, day of week, previous activity and activity length. The CASAS smart home is capable of recognising more complex activities due to the number of movement sensors employed, but isn't capable of recognition with multiple interacting inhabitants, or discrimination of activities with similar movement patterns yet different localised interactions. The CASAS smart home also lacks the fidelity of the WTH smart home which captures information from doors, windows, taps and appliances to reinforce the spatial information in determining the activity type.

The Institute for Multi-Sensor Processing and Content Analysis (IMPCA)[1] smart home environment is outlined in Section 2.6 and is used in this study. This environment includes non-invasive multi-camera video sensors to allow accurate tracking of multiple simultaneous inhabitants over several rooms, with contact sensors for doors and cupboards, pressure sensors on entries to doors, and appliance sensors to denote whether an appliance has changed state. The richness of the spatial and sensor data allows capturing of sufficient information to maximise activity disambiguation.

2.3 Human Activity Recognition Approaches

Activity recognition is the task of identifying an activity, defined earlier as an action or series of actions taken in pursuit of an objective. This view of activity recognition differs from others where authors classify the detection of walking, running or similar

[1] http://impca.curtin.edu.au

primitive actions as activity recognition. The confusion in activity and action definitions is also noted in reviews by (Aggarwal and Cai, 1999; Aggarwal and Park, 2004; Moeslund *et al.*, 2006; Turaga *et al.*, 2008; Aggarwal and Ryoo, 2011). In a smart home setting, objectives could include but not be limited to having breakfast, reading a newspaper, watching television, cooking, having a shower or going to sleep. The series of actions or events that one would need to recognise in order to determine whether an objective has been accomplished may involve walking to a table, opening a cupboard, sitting on a chair or even lying on a bed. The focus of this thesis is on the spatial component of activities as they can be obtained non-invasively from video tracking systems and for the majority of activities spatial signatures are unique (Chen *et al.*, 2005).

Video surveillance and automatic recognition of human activities is becoming increasingly important in modern society. Such recognition systems can be applied to problem domains such as security surveillance, aged care monitoring, behavioural biometrics, health care monitoring, sports analysis, human computer interaction (HCI) and smart homes (Dilger, 1997; Aggarwal and Cai, 1999; Ehlert, 2003; Aggarwal and Ryoo, 2011). Due to the diverse nature of the field of activity recognition, many attempts have been made to classify the existing approaches in to logical taxonomies. One popular approach is the top-down or bottom-up taxonomy. Top-down approaches recognise complex, semantically rich activities and principally have involved plan recognition techniques and hierarchical designs. The top-down taxonomy is analogous to the complex activity recognition and hierarchical taxonomies of Aggarwal and Ryoo (2011) and Turaga *et al.* (2008) respectively. Some examples of top-down approaches used in the modelling of high level behaviours include Dynamic Bayesian Networks (DBNs) (Intille and Bobick, 1998; Hamid *et al.*, 2003), abstract HMM (AHMM) (Nguyen *et al.*, 2002), multi-level HMM (Wojek *et al.*, 2006), coupled HMM (CHMM) (Oliver *et al.*, 2000), switching hidden semi-Markov model (S-HSMM) (Duong *et al.*, 2005) and stochastic grammars (Bobick and Ivanov, 1998; Ivanov and Bobick, 2000a; Pynadath and Wellman, 2000). Activity recognition of multi-tasked activities, such as those that occur concurrently or are interwoven, are primarily modelled as top-down approaches with bottom-up techniques used for low-level activity recognition (Ivanov and Bobick, 2000b; Kim *et al.*, 2010). Bottom-up approaches are flat in design and use low-level data such as spatial trajectories to develop simple models of activities. The bottom-up approaches are synonymous to the single-layered approaches in Aggarwal and Ryoo (2011), and the non-parametric actions of Turaga *et al.* (2008). Bottom-up approaches include template matching techniques (Bobick and Ivanov, 2001; Vlachos *et al.*, 2002b; Vaswani *et al.*, 2003; Chen *et al.*, 2004;

Chen and Ng, 2004; Chen *et al.*, 2005) and probabilistic models such as the HMM (Rabiner, 1989), and extensions to the HMM (Tan and Silva, 2003). Template matching techniques are typically more sensitive to noise in observations and variation in patterns of the same activity. Bottom-up probabilistic approaches such as Yamato *et al.* (1992) are able to better deal with uncertainty, yet do not scale well with large training sets and long activity sequences. The following sections expand on common template matching techniques such as DTW and Edit Distance variants, as well as the HMM as applied to activity recognition.

2.3.1 Template Matching Techniques

Template matching approaches use feature extraction techniques to derive sequence templates or exemplars for comparison to input sequences. Each activity class can have one or more templates, with classification decisions calculated by measuring the similarity of each template and an input sequence, and finding the activity class with the highest similarity. For robust activity recognition, template matching approaches must capture the variability in the templates for generalised approaches or employ a similarity matching technique that compensates for the observed variability or warping. Early attempts of quantifying the similarity of spatial templates used Euclidean distance metrics; however, these approaches were highly sensitive to temporal axis distortion as shown in Fig. 2.1)(Ratanamahatana and Keogh, 2004b; Chen *et al.*, 2005). DTW (Vlachos *et al.*,

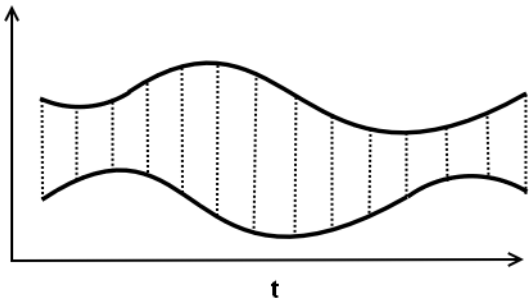


Figure 2.1: Euclidean distance with temporal warping.

2002c) and the Edit distance approaches of Chen and Ng (2004) and Chen *et al.* (2005) can compensate for the variability and warping between sequences for improved spatial template recognition. The authors from Chen *et al.* (2004) have also applied their

Edit Distance-based techniques to movement pattern strings, extracted from spatial sequences, to further improve the recognition performance of the approach.

2.3.1.1 Dynamic Time Warping (DTW)

DTW was formulated in Sakoe and Chiba (1978) to address temporal distortion via a process of non-linear time normalisation (see Fig. 2.2). Due to this time normalisation

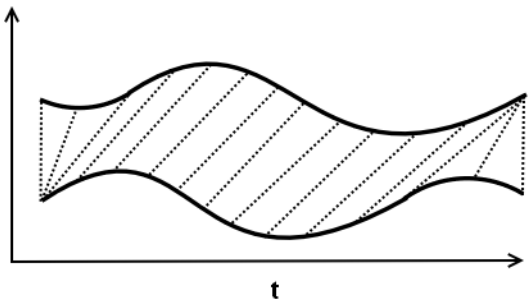


Figure 2.2: DTW with temporal warping.

property DTW has been applied in areas such as speech recognition (Sakoe and Chiba, 1978; Rabiner *et al.*, 1978; Das, 1982), trajectory recognition (Vlachos *et al.*, 2002c), bioinformatics (Aach and Church, 2001) and word image recognition (Rath and Manmatha, 2003). DTW has had varying success and is well known to be sensitive to noise and outliers, as a result of all sequence elements requiring mapping to a corresponding element(s) of an opposing sequence. There have been many DTW-based approaches that address warping problem domains or improve the robustness and/or runtime complexity of the approach. Some of these include derivative DTW (Keogh and Pazzani, 2001), time-warped longest common subsequence (T-WLCS) (Guo and Siegelmann, 2004), and iterative deepening DTW (Chu *et al.*, 2002).

The DTW algorithm provides elastic matching of two time series sequences $\mathbf{a}$ and $\mathbf{b}$ by minimising the cumulative distance between the sequences. In spatial activity recognition Euclidean distance $d(a_i, b_j)$ is commonly used as the local distance measure of individual trajectories a_i and b_j. DTW produces a warping path $\mathbf{w}$; a mapping between two sequences $\mathbf{a}$ and $\mathbf{b}$, where $\mathbf{w} = w_1, w_2, ..., w_\mathbf{k}$ for $max(|\mathbf{a}|, |\mathbf{b}|) \leq \mathbf{k} \leq |\mathbf{a}| + |\mathbf{b}| - 1$ ($w \equiv (i, j)$). A warping path $\mathbf{w}$ is constrained to ensure monotonicity and continuity.

For instance, a boundary constraint limits the warping path $\mathbf{w}$ such that the DTW calculation commences and finishes in diagonally opposite corners of the DP matrix, that is commences at $w_1 = (0, 0)$ and finishes at $w_k = (|\mathbf{a}| - 1, |\mathbf{b}| - 1)$. The monotonicity constraint ensures the points in the warping path are monotonically spaced in time, that is for $w_{k+1} = (i, j)$ then $w_k = (\hat{i}, \hat{j})$, where $i - \hat{i} \geq 0$ and $j - \hat{j} \geq 0$. The continuity constraint restricts the allowable steps in the warping path to adjacent cells in the DP matrix. It is stated formally as, given $w_{k+1} = (i, j)$ then $w_k = (\hat{i}, \hat{j})$, where $i - \hat{i} \leq 1$ and $j - \hat{j} \leq 1$.

To efficiently calculate the DTW distance of two time series sequences $\mathbf{a}$ and $\mathbf{b}$ a DP approach is utilised. A DP matrix C of size $|\mathbf{a}| \times |\mathbf{b}|$ is initialised according to (2.1). To calculate the DTW distance one applies (2.2) for values of $i = 2, \ldots, |\mathbf{a}|$ and $j = 2, \ldots, |\mathbf{b}|$. The resulting DTW distance is obtained from the DP matrix at $C(|\mathbf{a}|, |\mathbf{b}|)$.

$$\begin{aligned} C(1,1) &= d(a_1, b_1) \\ C(i,1) &= C(i-1,1) + d(a_i, b_1) \qquad i = 1, 2, \ldots, |\mathbf{a}| - 1 \\ C(1,j) &= C(1,j-1) + d(a_1, b_j) \qquad j = 1, 2, \ldots, |\mathbf{b}| - 1 \end{aligned} \tag{2.1}$$

$$C(i,j) = \min \left\{ \begin{array}{l} C(i-1,j-1) \\ C(i-1,j) \\ C(i,j-1) \end{array} \right\} + d(a_i, b_j) \tag{2.2}$$

A warping path or alignment can be recovered from C using a traceback procedure, originating at $C(\mathbf{a} - 1, \mathbf{b} - 1)$ and terminating at $C(1, 1)$, or by using a pointers matrix and retaining pointers to the local minima selected at each i and j during the DP calculation. The DTW algorithm provided in 2.2 has no local continuity constraints, thus not restricting the slope of the warping, unlike the approach of Sakoe and Chiba (1978). Typically, choosing an optimal local constraint is application and domain specific.

Calculation of DTW between two sequences is computationally expensive, $O(|a||b|)$. Runtime can be significantly reduced using DP banding techniques to minimise calculation of the DP matrix. Common banding techniques include use of the Sakoe-Chiba band (Sakoe and Chiba, 1978), Itakura parallelogram (Itakura, 1975), band DP constraint Das (1982) and Ratanamahatana-Keogh (R-K) band (Ratanamahatana and Keogh, 2004b). Use of both the Sakoe-Chiba and Itakura constraints are domain specific, whilst the band DP and R-K constraints are able to be used for most time series data. It was widely believed that wider bands result in improved recognition performance; however,

Ratanamahatana and Keogh (2004b) showed wider bands do not always result in optimal recognition rates and that the width and shape of the band also influence recognition performance.

2.3.1.2 Edit Distance on Real Sequence (EDR)

Edit Distance on Real Sequence (EDR) (Chen *et al.*, 2005) addresses the issue of local time shifting with spatial sequences similar to DTW, and improves robustness to noise in comparison to techniques such as Euclidean distance, DTW and LCSS. It is based on the Edit Distance approach from the string domain and bioinformatics, but is non-metric unlike Edit Distance. EDR calculates the distance between two strings via finding the minimum number of insertions, deletions and substitutions required to make two sequences identical. The EDR is calculated for two time series sequences **a** and **b** using a DP matrix C of size $(|\mathbf{a}|+1) \times (|\mathbf{b}|+1)$in conjunction with a matching threshold θ, Euclidean distance function $d(a_i, b_j)$, and iterating over i and j as shown in (2.3)-(2.6). The resulting EDR can be found at position $(|\mathbf{a}|+1, |\mathbf{b}|+1)$ within C.

$$C(0,0) = 0 \tag{2.3}$$

$$C(i,0) = i, \quad 1 \leq i \leq |\mathbf{a}| \tag{2.4}$$

$$C(0,j) = j, \quad 1 \leq j \leq |\mathbf{b}| \tag{2.5}$$

$$\begin{aligned} C(i,j) &= \min\{C(i-1,j-1) + dist_{EDR}(a_i,b_j), \\ &\qquad C(i-1,j)+1, \\ &\qquad C(i,j-1)+1\} \\ \text{where, } dist_{EDR}(a_i,b_j) &= \begin{cases} 0 & d(a_i,b_j) < \theta \\ 1 & d(a_i,b_j) \geq \theta \end{cases} \end{aligned} \tag{2.6}$$

A warping path or alignment can be recovered from C using a traceback procedure or by using a pointers matrix as per DTW.

Localised time shifting is achieved with EDR as the approach introduces gaps in sequences that correspond to changes in matched trajectories. The affine gap penalty and matching threshold θ also minimise the inclusion of noise in the distance calculation by

quantising the actual distance between spatial elements as linear increasing values for gaps and zero for matches. The affine gap penalty also allows EDR to outperform LCSS, which provides no penalty for gaps or gap length. In Chen *et al.* (2004), EDR is applied to movement pattern strings (MPS's) resulting in the Edit Distance on MPS (EDM) approach that has reduced dimensionality. To further reduce computational cost of EDM, MPS's are transformed into frequency vectors and a modified frequency distance is applied (as frequency distance between two frequency vectors is an approximation of the Edit Distance (Kahveci and Singh, 2001)).

2.3.1.3 Edit Distance with Real Penalty (ERP)

Edit Distance with Real Penalty (ERP) (Chen and Ng, 2004) is a metric distance function based on EDR and Edit distance. ERP avoids the use of θ-based thresholding in the distance calculation (unlike EDR), and does not replicate previous elements in gaps (unlike DTW) in order to satisfy the triangle inequality of metrics. This allows ERP to be used with indexing structures for efficient retrieval tasks. ERP calculates the distance between two strings similar to EDR except an Euclidean distance penalty is applied for non-gap elements and a constant value is given for gap distances.

For two time series sequences $\mathbf{a}$ and $\mathbf{b}$, ERP is derived using a DP matrix C of size $(|\mathbf{a}|+1) \times (|\mathbf{b}|+1)$ whilst iterating over i and j as shown in (2.7)-(2.10). The resulting EDR metric can be found at position $(|\mathbf{a}|+1, |\mathbf{b}|+1)$ within C.

$$C(0,0) = 0 \tag{2.7}$$

$$C(i,0) = a_i + a_{i-1}, \quad 1 \leq i \leq |\mathbf{a}| \tag{2.8}$$

$$C(0,j) = b_j + b_{j-1}, \quad 1 \leq j \leq |\mathbf{b}| \tag{2.9}$$

$$\begin{aligned} C(i,j) = \min \{ & C(i-1,j-1) + dist_{ERP}(a_i, b_j), \\ & C(i-1,j) + dist_{ERP}(a_i, gap), \\ & C(i,j-1) + dist_{ERP}(gap, b_j) \} \end{aligned} \tag{2.10}$$

$$\text{where, } dist_{ERP}(a_i, b_j) = \begin{cases} |a_i - b_j| & a_i, b_j \text{ not a gap} \\ |a_i| & b_j \text{ is a gap} \\ |b_j| & a_i \text{ is a gap} \end{cases}$$

Warping paths or alignments can be derived as per other DP-based approaches such as DTW and ERP. Experimental validation by Chen and Ng (2004) using spatial data from different problem domains demonstrates that ERP has strong discriminatory characteristics, performing marginally better in classification than both DTW and ERP with simulated and captured data sets. As ERP quantifies all variability in its matching criteria, it is more sensitive to noise than EDR.

2.3.2 Hidden Markov Models

The Hidden Markov Model (HMM) is a stochastic state transition model (assumes Markovian dynamics), capable of dealing with time sequential data (Rabiner, 1989). It was first applied in the activity recognition domain by Yamato *et al.* (1992), where mesh features were extracted from time sequential images of tennis strokes and used in training and evaluation of a discrete model. Since then the HMM has been utilised extensively in activity recognition research, particularly through the multi-layer and hierarchical forms (Oliver *et al.*, 2000; Bui *et al.*, 2001; Duong *et al.*, 2005; Truyen *et al.*, 2005; Wojek *et al.*, 2006). Adoption of the HMM has been motivated in this thesis by the models ability to deal with noisy observations and its' high discriminatory properties.

A discrete HMM is characterised by a number of hidden states N, distinct observation symbols per state M, state transition probability matrix A ($A = \{a_{ij}\}$), observation symbol probability distribution matrix B ($B = \{b_j(k)\}$) and the initial state distribution vector π. A derived HMM λ is typically represented by the tri-tuple of parameters $\{\pi, A, B\}$, which represent the following:

$$\begin{aligned} \pi &= Pr(q_1 = S_j), \quad 1 \leq i \leq N \\ a_{ij} &= Pr(q_{t+1} = S_j | q_t = S_i), \quad 1 \leq i, j \leq N \\ b_j(k) &= Pr(v_k \text{ at } t | q_t = S_j), \quad 1 \leq j \leq N, 1 \leq k \leq M \end{aligned}$$

where q_t is the state at time t, S is the individual states such that $S = \{S_1, S_2, \ldots, S_N\}$ and V denotes the individual symbols $V = \{v_1, v_2, \ldots, v_M\}$. The HMM model parameters π, A, B are derived using the Baum-Welch (Forward-Backward) algorithm; however, scaling Rabiner (1989) is required in both the model estimation and inferencing, due to the lengthy observation sequences. The probability of an observation sequence $\mathcal{O}$ given

the derived model λ ($Pr(\mathcal{O}|\lambda)$) is calculated in later experiments using the forward algorithm with scaling.

The discrete HMM can produce good discriminative models; however, the following needs to be considered when applying the approach to a spatial recognition domain (due to the inherent noise and long length of sequences). The issue of choosing an appropriate number of hidden states N for each model is important, as the value of N affects both the runtime of the Baum-Welch algorithm and the discriminatory capability of the model. If N is too small, the training and inferencing runtime are smaller but the models ability to discriminate effectively also decreases. In contrast if N is large, training and inferencing complexity is increased along with discrimination capability; however, the model may overfit the data if N is too large. Also, if the spatial sequence length is long (> 1000) the HMM may also fail to adequately represent dissimilarity for an activity and across the training data (particularly with small environments), resulting in decreases in disambiguation and classification accuracy. Importantly, HMMs do require large data sets for training of models and for good discrimination. If only limited spatial sequences are available for training, the model will not be able to generalise appropriately and will suffer from lower recognition accuracy when evaluated with observed sequences. In addition to the amount of training data required, the discrete HMM also encounters issues when calculating the $Pr(O|\lambda)$, where O is the observed spatial sequence, if an observed sequence contains a symbol not present in the training sequences (can arise with noise). In this case, ($b_j(k) = 0$) the derived probability of the overall sequence, calculated using a scaled forward procedure will tend to zero. This can be minimised by initialising the symbol probability matrix B to small positive values; however, the low probability of the observed symbol will still dramatically reduce the overall sequence probability.

2.4 Bioinformatics: Sequence Alignment

The field of bioinformatics is concerned with the collection, classification, storage, and analysis of biochemical, biological and genetic information. Much of the research in this field focuses on finding genes, modelling evolution, protein structure prediction and DNA, RNA or protein sequence alignment (Waterman, 1995). Biological scientists have been using bioinformatics, in particular sequence similarity or alignment approaches, to

classify unknown DNA[2], RNA[3] and protein sequences[4] obtained through laboratory research. Furthermore the approaches have allowed them to develop phylogenies describing the evolutionary history of an organism or gene[5].

In practical pattern-matching applications, including searching genomes for similar DNA sequences and activity recognition, exact matching is not always pertinent. Generally, it is more important to find approximate matches to a given pattern thus allowing for innate pattern variability (Crochemore and Rytter, 2002). Sequence similarity has both quantitative and qualitative aspects. Quantitatively, similarity measures produce values which describe the degree of similarity of two sequences. Alignments, which are mutual arrangements of two sequences (Refer to Fig. 2.3), give a qualitative answer as they show where two sequences are similar and where they differ. Optimal alignments, which exhibit maximum correspondence of two sequences with the least differences, can be used to determine similarity quantitatively as well as qualitatively. Thus, it is possible to use optimal sequence alignments to measure the degree of similarity and to also find regions of dissimilarity (Waterman, 1995).

```
2 3 3 3 4 1 - - 4 2 2 1 1
|       | |     | | |   |
2 2 2 - 4 1 3 3 4 2 2 2 1
```

Figure 2.3: An example alignment of two short sequences. The symbol - refers to an *indel* and represents an insertion or deletion in either sequence. Two gaps are present in the alignment; one of size 1 and the second of size 2.

Sequence similarity in bioinformatics is quantified using either similarity or distance approaches. Similarity approaches use a function that associates a numeric value with a pair of sequences such that higher values indicate greater similarity. Examples of such similarity approaches include the Needleman-Wunsch global alignment (Needleman and Wunsch, 1970) and Smith-Waterman(SW) local alignment (Smith and Waterman, 1981). Distance approaches are similar; however, they treat sequences as points in a metric space. Informally, distance approaches include a function that associates a numeric value with a pair of sequences, with high values indicating high degrees of dissimilarity.

[2]Deoxyribonucleic Acid (DNA) sequences consist of four building blocks: adenine (A), thymine (T), guanine (G) and cytosine (C)

[3]Ribonucleic Acid (RNA) sequences consist of four building blocks: adenine (A), uracil (U), guanine (G) and cytosine (C)

[4]Protein sequences consist of amino acids or peptides

[5]A gene is a segment or sequence of DNA that codes a cellular product

Unlike similarity approaches, distance approaches satisfy the mathematical axioms of a metric (Waterman, 1995).

The sequence alignment problem can be formally stated as follows: given two query sequences **a** and **b** with symbols a_i and b_j for $i = 1, ..., |a|$ and $j = 1, 2, ..., |b|$, find the best matching alignment in relation to the specified optimisation criteria, maximisation of similarity or minimisation of distance. To derive optimal alignments, each symbol a_i is compared sequentially with the symbols b_j of the other sequence. During this stage the local similarity or distance is calculated between the opposing symbols and using techniques such as dynamic programming (DP), optimal subalignments resulting in an optimal alignment can be determined. With the maximising similarity criteria a positive score is associated with matching symbols, while negative scores are given to non-matching symbols and insertions/deletions (referred to as indels). Indels, denoted by the symbol (-) in an alignment diagram, are used to represent insertions or deletions in either sequence and are incorporated into alignments to fill gaps caused by differences in either of the sequences. For non-matching symbols, the choice whether to include an indel in the alignment or not is dependent on which of the options is more optimal, that is has a higher total similarity or smaller overall distance.

An alignment assumes that two sequences **a** and **b** satisfy the following constraints (Eidhammer *et al.*, 2004):

1. All symbols in the sequences **a** and **b** must also be in the alignment. For example, if **a** = [123] and **b** = [124] the resulting alignment must be of the form:

   ```
   - 1 - 2 - 3 -
   |   |
   - 1 - 2 - 4 -
   ```

   where - represents zero or more indels.

2. All symbols in the alignment must appear in the same order as defined in the sequences **a** and **b**, except that zero or more indels or gaps may be present between the symbols, as in the above example.

3. Symbols of one sequence can be aligned with an indel in the other sequence. For example, if **a** = [123] and **b** = [13], the 2 in **a** can be aligned with an indel as shown below:

```
1 2 3
|   |
1 - 3
```

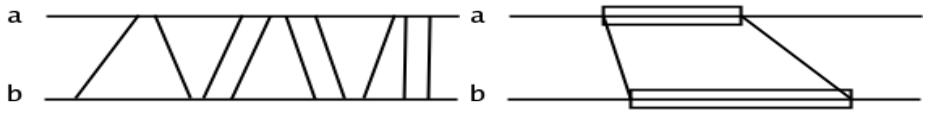
4. Indels of different sequences *cannot* be aligned together.

An example pairwise alignment of one dimensional sequences **a** = [12223421144]and **b** = [111341121114] is shown in Fig. 2.4. The given alignment contains seven matching symbols, three symbol mismatches and three indels.

```
1 2 2 2 3 4 - - 2 1 1 4 4
|       | |     | | |   |
1 1 1 - 3 4 1 1 2 1 1 1 4
```

Figure 2.4: An example sequence alignment containing gaps of size one and two.

Sequence alignment was pioneered in bioinformatics with a DP approach in Needleman and Wunsch (1970). This technique aligned two protein sequences across their entirety, maximising the similarity score of the matching individual protein constituents (amino acids). A DP basis is used as it is possible for optimal alignments to be calculated from incrementally derived subalignments as each subalignment is itself optimal. Local alignment approaches such as SW Smith and Waterman (1981) differ from global techniques as they find and quantify related regions of similarity *within* sequences. The local alignments are typically found via an optimsation search process, originating at the beginning of the sequences until the ends. An illustrative example of global and local alignment using sequences **a** and **b** can be found in Fig. 2.5. Recently, sequence



(a) Global Alignment (b) Local Alignment

Figure 2.5: Schematic global and local alignments between two sequences.

alignment techniques have also been applied in pattern recognition approaches such as speech recognition, string matching and matching moving object trajectories from video surveillance data (Vlachos *et al.*, 2002b,a; Chen *et al.*, 2004; Chen and Ng, 2004; Chen *et al.*, 2005).

Sequence alignment approaches are applied in bioinformatics to derive the relationship of unknown biological sequences in regards to known, empirically well understood and genetically mapped genes, sequences or proteins. To discriminate biologically meaningful alignments from ones due to chance, a substitution matrix is sometimes applied to the substitution score calculation of an alignment. A substitution matrix describes the rate or probability that an element in a sequence will transition to another and is derived from iterative pairwise sequence alignment calculations. In problem domains such as biology and speech recognition, not all symbols will have equal probability of transitioning to others. Two popular bioinformatics substitution matrices are the Point Accepted Mutation (PAM) and Blocks Substitution (BLOSUM) matrices (Apostolico and Giancarlo, 1998; Barton, 1998). The PAM matrix is used for protein elements (referred to as amino acids) and describes a fixed probability that a particular symbol or amino acid will change to another symbol (inclusion of reversion) according to an evolutionary protein model. They key issue associated with the PAM matrix is that its weights are calculated based on closely related sequences. As a result PAM weights are inferred on all sequences, including those of distantly related sequences, which may result in poor similarity values and alignments for related sequences. The BLOSUM matrix addresses some of the issues of the PAM matrix in relation to aligning distant sequences. A BLOSUM matrix is derived by applying blocks to multiple sequence alignments of evolutionary divergent proteins. Each block represents a cluster of sequences with a defined percentage similarity, the most common being the BLOSUM 62 matrix. In Prlic *et al.* (2000), the authors demonstrated that sequence alignment accuracy is highly dependent on the substitution matrix employed, particularly with distantly related sequences. It is possible to learn a substitution matrix for each activity from its resulting alignments; however, it is likely that such an approach will still be subject to Markovian constraints and its limitations as per the HMM.

Pairwise sequence alignment techniques such as the ones discussed in this chapter can employ different gap penalty functions to penalise gap lengths (resulting from insertions or deletions). The most popular is the affine gap penalty function, defined as $g(x)$ for gap length x, which can allow a gap open penalty ϕ and a gap extension penalty ψ to be applied according to (2.11) (Sankoff and Kruskal, 1999b).

$$g(x) = \phi + \psi(x) \tag{2.11}$$

If $\phi = 0$ the affine gap penalty model approximates the linear gap model used with Edit

Distance-based approaches and local alignment techniques.

2.4.1 Longest Common Subsequence (LCSS)

Like its name suggests LCSS finds the longest common subsequence from two sequences, where subsequences need not be adjacent. The LCSS is calculated using a DP matrix C (depicted in (2.12)) of size $(|\mathbf{a}|+1)\times(|\mathbf{b}|+1)$, through the application of (2.13) for $i = 1, \ldots, |\mathbf{a}|$ and $j = 1, \ldots, |\mathbf{b}|$.

$$\begin{bmatrix} C(0,0) & \ldots & C(0,|b|) \\ \vdots & \ddots & \vdots \\ C(|a|,0) & \ldots & C(|a|,|b|) \end{bmatrix} \tag{2.12}$$

$$C(i,j) = \begin{cases} 0 & i = 0 \text{ or } j = 0 \\ C(i-1,j-1)+1 & i,j \geq 1 \text{ and } a_i = b_j \\ \max\{C(i,j-1), C(i-1,j)\} & i,j \geq 1 \text{ and } a_i \neq b_j \end{cases} \tag{2.13}$$

LCSS length is used as a similarity score for the alignment and is obtained from $C(|a|,|b|)$. To determine the resulting alignment from LCSS one can utilise a second matrix of size $(|\mathbf{a}| \times |\mathbf{b}|)$ to retain pointers to which of the three choices $(C(i-1,j-1), C(i-1,j)$ or $C(i,j-1))$ were chosen at each suboptimal solution. Inadvertently, a traceback procedure may also be used to reconstruct the optimal path from $C(|\mathbf{a}|,|\mathbf{b}|)$ to $C(1,1)$ by determining which of the previous terms $C(i-1,j-1), C(i-1,j)$ or $C(i,j-1)$ resulted in $C(i,j)$. LCSS achieves robust recognition in the presence of noise by ignoring regions of dissimilarity and maximising element matching between sequences. A similar LCSS based approach has been proposed by Vlachos *et al.* (2002a), in which the authors incorporated a constant δ to restrict how far points can match in time and a matching threshold ϵ that specified how close trajectories must be in order to match. The resulting similarity value from the LCSS variant was then normalised using the minimum of the sequence lengths.

LCSS achieves its robust recognition in the presence of noise by ignoring regions of dissimilarity and maximising element matching between sequences. In activity recognition,

one wishes to accurately recognise activities through the intrinsic noise incorporated during video tracking. By ignoring all dissimilarity, LCSS is subject to misclassification with similar sequences from different classes. Normalisation of the LCSS algorithm in Vlachos *et al.* (2002a) does provide a coarse approximation of dissimilarity; however a more accurate means of dissimilarity quantification is needed.

2.4.2 Edit Distance

Edit Distance (or otherwise known as Levenshtein distance) is a popular global alignment metric used in bioinformatics for sequence alignment and distance measurement. It has also seen significant use over the years in diverse pattern recognition applications due to its strong discriminatory properties. Edit Distance functions via finding the minimum number of edit operations comprising insertions, deletions and substitutions, to transform one sequence into another. A penalty of one is associated with each edit operation; however, more complex penalty schemas involving substitution matrices and gap penalty models have been applied in bioinformatics and pattern recognition applications. As a metric, Edit Distance satisfies the metric axioms below and can be more efficiently applied to retrieval tasks than non-metric distance measures (Sankoff and Kruskal, 1999c):

Non-negative property $d(\mathbf{a}, \mathbf{b}) \geq 0$, for all $\mathbf{a}$ and $\mathbf{b}$.

Zero property $d(\mathbf{a}, \mathbf{b}) = 0$, if and only if $\mathbf{a} = \mathbf{b}$.

Symmetry $d(\mathbf{a}, \mathbf{b}) = d(\mathbf{b}, \mathbf{a})$, for all $\mathbf{a}$ and $\mathbf{b}$.

Triangle inequality $d(\mathbf{a}, \mathbf{b}) + d(\mathbf{b}, \mathbf{c}) \geq d(\mathbf{a}, \mathbf{c})$, for all $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$.

The Edit Distance can be calculated for two symbolic sequences $\mathbf{a}$ and $\mathbf{b}$ using a DP matrix C of size $(|\mathbf{a}|+1) \times (|\mathbf{b}|+1)$ and iterating over i and j as shown in (2.14)-(2.17). The resulting Edit Distance can be found at position $(|\mathbf{a}|+1, |\mathbf{b}|+1)$ in C.

$$C(0,0) = 0 \tag{2.14}$$

$$C(i,0) = i, \quad 1 \leq i \leq |\mathbf{a}| \tag{2.15}$$

$$C(0,j) = j, \quad 1 \leq j \leq |\mathbf{b}| \tag{2.16}$$

$$
\begin{aligned}
C(i,j) &= \min\{C(i-1,j-1) + dist_{ED}(a_i, b_j), \\
&\qquad C(i-1,j)+1, \\
&\qquad C(i,j-1)+1\} \qquad (2.17) \\
\text{where, } dist_{ED}(a_i,b_j) &= \begin{cases} 0 & a_i = b_j \\ 1 & a_i \neq b_j \end{cases}
\end{aligned}
$$

Sequence alignments can be recovered from C using procedures outlined previously.

Edit distance and its variants EDR (Chen *et al.*, 2005) and ERP (Chen and Ng, 2004) allow intrinsic warping of sequences due to the inclusion of gaps in the derivation of the optimal alignment. A key limitation of the Edit Distance approach in some domains is its linear gap penalty function, which provides linear weighting according to the gap length. This prevents one from applying more significant penalties with large gaps to prevent cases of extraneous warping.

2.4.3 Smith-Waterman Local Alignment

Smith-Waterman (SW) local alignment was originally developed in (Smith and Waterman, 1981) to locate biological sequence patterns within known sequence databases. SW is similar to the Needleman-Wunsch global alignment approach (Needleman and Wunsch, 1970), except that SW includes an extra zero. Inclusion of the zero allows termination of subsequence alignments that perform poorly, as non-matching subsequences produce negative similarity, which reduces the similarity between the subsequences. When the similarity becomes negative, the zero of the SW relation terminates any further decrease in similarity and allows new optimal subsequences, referred to as *local* alignments, to be found. SW functions by maximising the similarity score S between two segments of the sequences $\mathbf{a} = a_i \cdots a_k$ and $\mathbf{b} = b_j \cdots b_l$ given a match cost α, mismatch penalty δ and an indel penalty γ, where $1 \leq i \leq k \leq n$ and $1 \leq j \leq l \leq m$. This is formally shown in (2.18).

$$
S_{opt}(a,b) = \max_{\substack{1 \leq i \leq k \leq n \\ 1 \leq j \leq l \leq m}} \{S_{\delta,\gamma}\{a_i \cdots a_k, b_j \cdots b_l\}\} \qquad (2.18)
$$

where $S\{a_i \cdots a_k, b_j \cdots b_l\} = \alpha u + \delta v + \gamma w$ and u, v, w are the numbers of matches, mismatches and gaps respectively.

Gap scores resulting from indels are typically a function of the gap length l, denoted by $g(l)$. A linear gap model is employed in SW local alignment, where $g(l) = -l\gamma$, assigning equal weight for gaps. In this linear gap model it is assumed that the probability of a gap occurring in a sequence is the same anywhere along the sequence.

To calculate the SW similarity of two sequences $\mathbf{a}$ and $\mathbf{b}$ (2.20)-(2.22) is applied to the DP matrix C of size $(|\mathbf{a}| + 1) \times (|\mathbf{b}| + 1)$ (2.19) for $i = 0, 1, \ldots, |\mathbf{a}|$ and $j = 0, 1, \ldots, |\mathbf{b}|$. The resulting SW similarity value is found by finding the maximum value in C.

$$\begin{bmatrix} C(0,0) & \ldots & C(0,|\mathbf{b}|) \\ \vdots & \ddots & \vdots \\ C(|\mathbf{a}|,0) & \ldots & C(|\mathbf{a}|,|\mathbf{b}|) \end{bmatrix} \tag{2.19}$$

At each $C(i, j)$ where $i, j \neq 0$, four choices (match or mismatch, gap in $\mathbf{a}$, gap in $\mathbf{b}$ or start a new subsequence) are evaluated with the choice corresponding to the maximum similarity value being selected for each $C(i, j)$. The match or mismatch score at each $C(i, j)$ is derived using $s(a_i, b_j)$, while the gap scores for the sequences are derived using the linear gap model. If a negative similarity score results from $C(i-1, j-1)+s(a_i, b_j)$, $C(i-1, j)+\gamma$ and $C(i, j-1)+\gamma$, due to poor subsequence correspondence, then the fourth option of starting a new subsequence, represented by zero, is selected as the maximum.

$$C(i,0) = 0, \quad 0 \leq i \leq |\mathbf{a}| \tag{2.20}$$

$$C(0,j) = 0, \quad 0 \leq j \leq |\mathbf{b}| \tag{2.21}$$

$$\begin{aligned} C(i,j) &= max\{C(i-1,j-1) + s(a_i,b_j), \\ &\qquad C(i-1,j) - \gamma, \\ &\qquad C(i,j-l) - \gamma,\ 0\} \end{aligned} \tag{2.22}$$

$$\text{where, } s(a_i,b_j) = \begin{cases} \alpha & a_i = b_j \\ \delta & a_i \neq b_j \end{cases}$$

Resulting SW alignments are dependent on the values of the gap penalty γ and match cost α. If γ is larger in relation to the average mismatch penalty δ, mismatches are favoured over gaps, producing shorter, more compact alignments. The opposite occurs when γ is smaller than the average mismatch penalty δ. If $\alpha \gg \gamma$ or the mismatch penalty δ, SW ignores mismatches and gaps and therefore behaves similar to LCSS. The

resulting complexity of the SW algorithm is $O(|\mathbf{a}||\mathbf{b}|)$, as completion of the DP matrix requires $|\mathbf{a}| \times |\mathbf{b}|$ steps.

2.5 Biologically-Inspired Computational Models

Biological systems are said to often resemble a fractal-the closer you look, the more detail that emerges Bhalla (2003). Biology been used as inspiration for many computational models including the artificial neural network (ANN) (Haykin, 1999), swarm intelligence approaches such as ant colony optimisation (ACO) (Dorigo *et al.*, 1999), evolutionary computing (Passino, 2002) and artificial immune systems (AIS) (de Castro and Timmis, 2002) to name a few. An ANN is modelled of the cerebral cortex of the brain and comprises a series of input, hidden and output layers with interconnected nodes (processing elements or neurons) that contain an activation function. The ANN receives input from the input layer, processes it in the hidden layers and then sends the result to the output layer similar to neural processing in the brain. The ACO approach models ant behaviour in response to food sources to solve combinatorial optimisation problems. Effectively each ant of a colony builds a solution to a given problem (i.e. how to get to a food source in the biological case) laying down a pheromone trail in doing so. Those ants that find a more optimal solution will return back to the colony with food more often than those that don't, resulting in increased strength of the pheromone trail, attracting further ants to the optimal case. Evolutionary computing is based on a model evolution that continually and incrementally redesigns the structure and parameters of organisms to maximise fitness and survival in a given environment. It is typically applied through genetic algorithms for solving global optimisation problems. Finally, AIS's are defined by de Castro and Timmis (2002) as "adaptive systems, inspired by theoretical immunology and observed immune functions, principles and models, which are applied to problem solving." AIS approaches typically apply immune algorithms (primarily negative selection) with affinity measures to determine "non-self" cases and are popular in fault and anomaly detection problems.

Biology serves as a useful basis for solving many real world problems due to the robustness, adaptiveness, diversity, error tolerant mechanisms and decentralised natures of its systems (Paton, 1994). (de Castro and Timmis, 2002) defines three different categories of approaches that have been used in biology and computing:

Biologically motivated computing Biological models are used as a sources of inspiration for the development of computational models, of which the ANN is a good example.

Computationally motivated biology Computing provides models and inspiration for biology. A simple example of this is the application of cellular automaton in Artificial Life.

Computing with biological mechanisms Information processing capabilities of biological systems (such as DNA) are used to replace or supplement computing systems. DNA-based computing is a prime example.

The approach taken in this thesis is the one of biologically motivated computing, where a robust biological process is modelled to address the issue of sequence variability and tracking system noise in activity recognition.

2.6 Activity Data Sets

Spatial activity sequences utilised in this research are captured in the IMPCA smart home environment located at Curtin University. The smart home consists of a series of rooms with several cameras per room with overlapping FOVs and contains common household items and appliances to allow capturing of everyday activities. Pressure, contact and switch sensors are also installed in the floor, furniture, cupboards and appliances, which are not utilised in this research. The layout of the smart home is provided in Fig 2.6 with an image of Room 1 provided in Fig. 2.7.

2.6.1 Camera Tracking System

Spatial activity sequences are captured for this research using the distributed, multi-camera, tracking system of Nguyen *et al.* (2002). Initial investigations focussed on the object tracking system of Peursum *et al.* (2003), which used multiple Kalman (and later particle) filters and cameras to estimate an objects position, but the system suffered from significant blob segmentation and sensitivity issues when applied to activity recognition

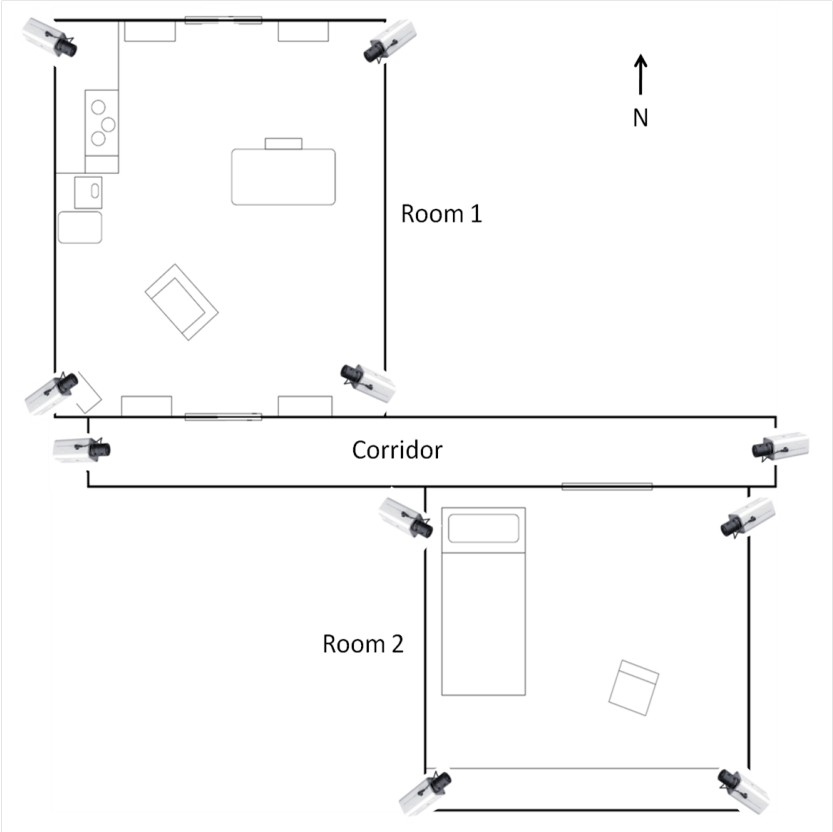


Figure 2.6: Layout of the Smart Home environment.

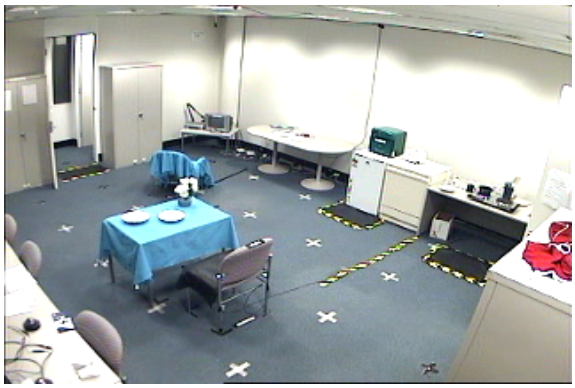


Figure 2.7: Room 1 of the Mock Smart House Environment.

contexts. To assist with explanation of the experimental methodology and results in later chapters, the system of Nguyen *et al.* (2002) is described here, with particular

relevance to the process involved with generating the activity sequences.

The system of Nguyen *et al.* (2002) utilises two modules for tracking of individuals: a camera processing module (CPM), which is directly connected to the camera and conducts blob segmentation, matching and aggregation, and a central module (CM) which manages the objects being tracked and assigns objects to CPM's for tracking. Each camera and corresponding CPM connects to the CM via the IP network backbone as shown in Fig. 2.8. Only one camera is used to track an object at any one time (rather than multiple cameras) for computational efficiency. The decision of which camera to choose is based on which has the best view of the subject, under the assumption that the larger the bounding box, and the closer the subject is to the camera, the better the view. Positional variability can be encountered with this approach if the CM assigns a new camera to track an object, even when properly calibrated.

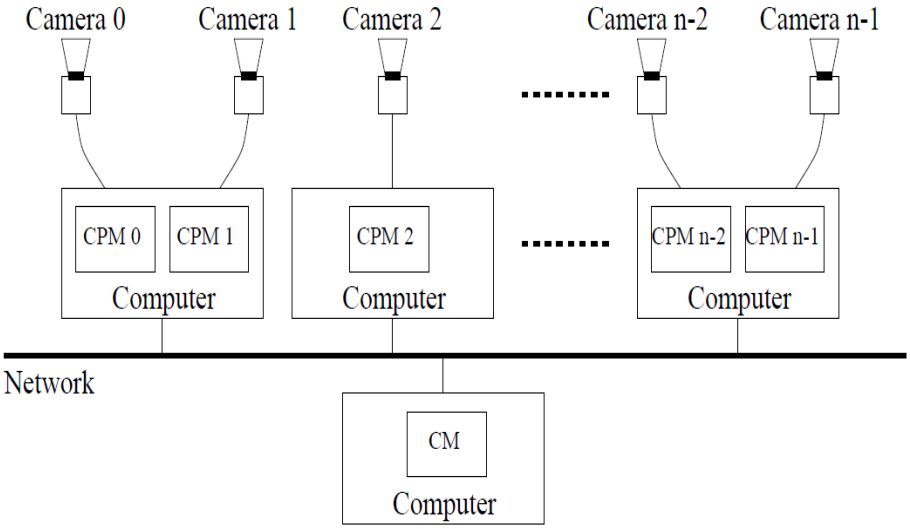


Figure 2.8: Configuration of the distributed camera tracking system (from Nguyen *et al.* (2002))

Each CPM in Fig. 2.8 identifies blobs (bounding boxes) corresponding to people in the smart home environment by dynamically modelling the background through averaging a series of previous images and applying an exponential decay function to update the background over time. This background subtraction approach is suitable for controlled environments such as a smart house as the background seldom changes; however,

with more dynamic environments approaches such as Gaussian Mixture Models (GMMs) would be more appropriate. Following the background modelling and initialisation, the foreground is then subtracted and the resulting bounding box of the foreground blob is calculated using the chain-code algorithm. Each bounding box or blob is assigned a position (x, y), size (w, h) and average colour (r, g, b) derived from its characteristics. Recursive blob merging is applied for blobs within a pre-specified distance due to intrinsic tracking noise and where object and background colours may be similar. Remaining small blobs are assumed to be noise and are removed. The output of this process is shown in Fig. 2.9. Blobs derived from a single camera view are matched to Kalman

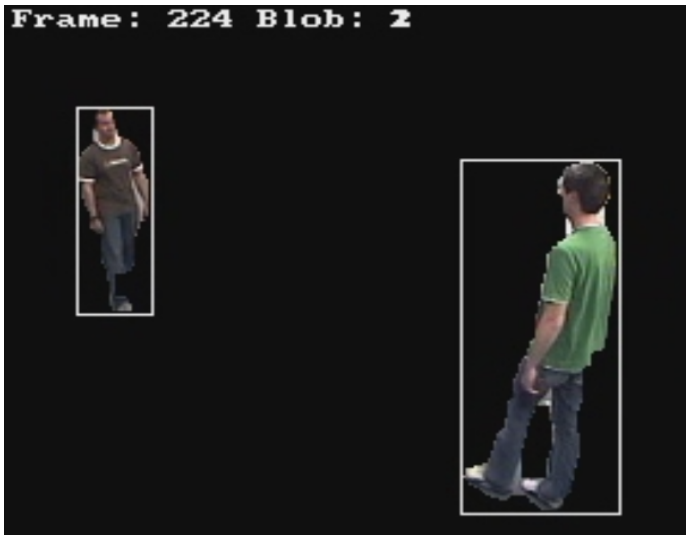


Figure 2.9: An example of 2 segmented blobs post-background subtraction and blob merging

filters by calculating the probabilistic distance of a blob vector and Kalman filter state vector, with vectors comprising blob position, size and average rgb colour. The system then finds the set of blob and Kalman filter state pairs that minimise the distance, within specified thresholds. Any remaining unmatched blobs are treated as lost objects; objects that were previously being tracked but haven't been observed in previous frames (possibly due to occlusion), and attempts to match these based on only average colour and blob size using the last estimate of the Kalman filter. Any remaining blobs are then passed to the CM for further testing against other CPMs. The overall process is outlined in Fig. 2.10.

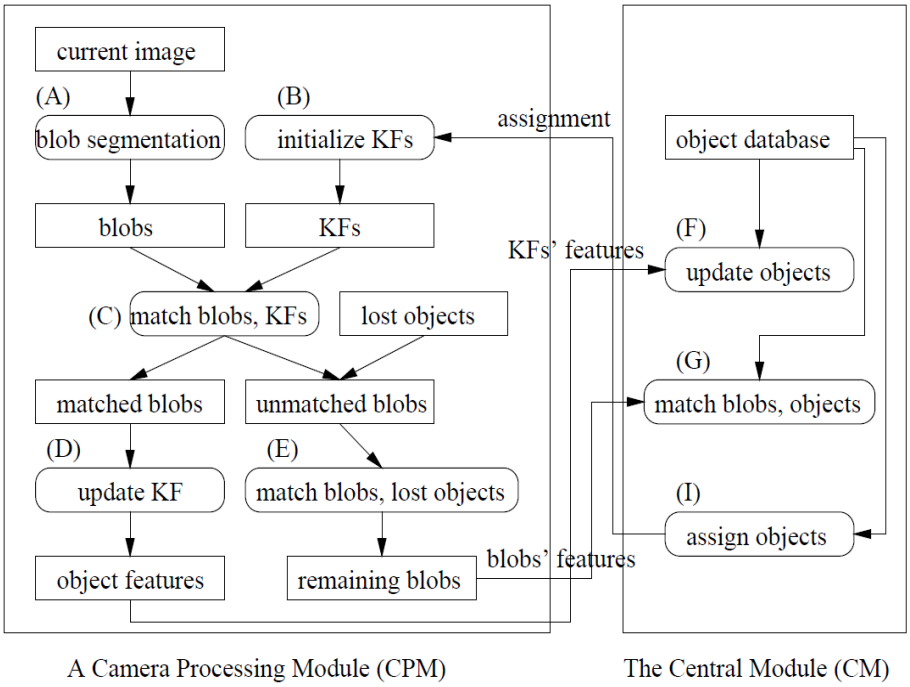


Figure 2.10: CPM and CM processing with the distributed camera tracking system (from Nguyen *et al.* (2002))

The output of the system is a set of relative x, y coordinates every time t and an object identifier, with the origin $0, 0$ set to the NW corner of the smart house environment. For each activity captured using this system, the object identifier corresponding to the individual conducting the activity is identified and used to filter the respective x, y coordinates. Each filtered sequence of x, y coordinates corresponding to the activity are then ground truthed against the video to identify tracking accuracy. Three datasets are captured using this system for experiments in Chapters 3 - 6 and are summarised in Table 2.1. Further details on the composition of the datasets used throughout this thesis are provided in the following sections.

2.6.2 Dataset A (10 activities)

The purpose of dataset A is to represent a diverse group of spatial activities that are captured a mock smart home environment. Activities range from 60 - 90 seconds in

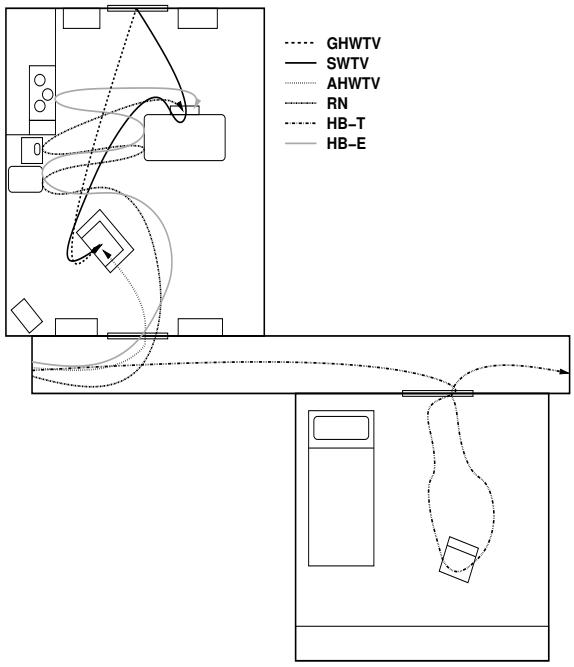
Table 2.1: Characteristics of datasets A,B and C used in this thesis.

Dataset	No. of Different Activities	No. of Sequences per Activity	Activity Description
A	10	20	Diverse Activities (Rooms 1 and 2)
B	3	20	Spatially similar activities (Room 1 only)
C	12	20	Diverse Activities (Room 1 only)

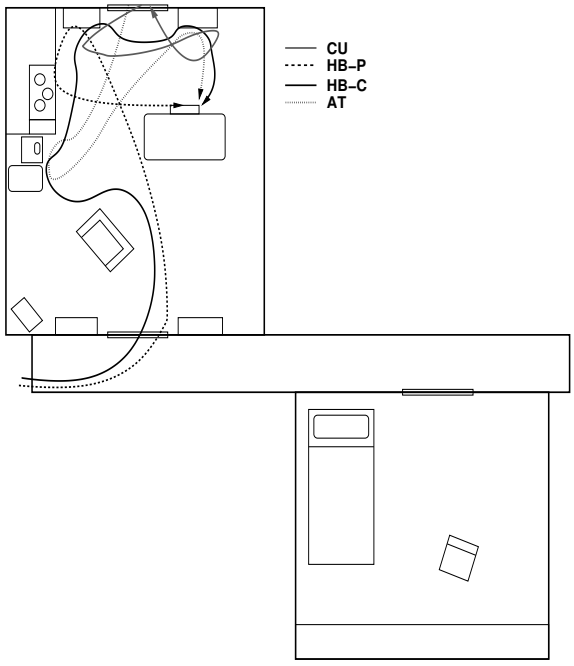
length, with subjects entering the smart home from the N or S doors of Room 1, or the N door of Room 2 and carrying out the specified activities in a consistent manner. Video is captured at 10 fps and processed by the video tracking system of Nguyen *et al.* (2002). The dataset comprises 10 single-person activities with the actions of each activity outlined in Table 2.2. Figure 2.11 outlines the spatial paths taken for the 10 activities.

Table 2.2: Dataset A activities and actions.

Activity Name	**Activity Description**
Get Home-Watch TV (GHWTV)	Enter through Room1 North door and sit on TV chair
Have a Snack-Watch TV (SWTV)	Enter through Room1 North door, sit down at dining table to eat snack and then move to sit on TV chair
At Home-Watch TV (AHWTV)	Enter through Room1 South door and sit down on TV chair
Read Newspaper (RN)	Enter down corridor, enter Room2, sit down on bed chair, read for a while, then leave
Have Breakfast-Toast (HB-T)	Enter though Room1 South door, go to fridge for OJ, put OJ on dining table, toast bread, eat at dining table
Have Breakfast-Eggs (HB-E)	Enter through Room1 South door, go to fridge for OJ, put OJ on dining table, cook eggs, eat at dining table
Clean Up (CU)	Enter through Room1 North door, and pickup items in kitchen, place some items in NW cupboard, then place remaining items in NE cupboard, leave through Room1 North door
Have Breakfast-Porridge (HB-P)	Enter through Room1 South door, get porridge from NW cupboard, heat up at stove in kitchen, eat at dining table
Have Breakfast-Cereal (HB-C)	Enter through Room1 South door, get milk from fridge, a bowl from NW cupboard, cereal from NE cupboard, then sit at dining table to eat
Afternoon Tea (AT)	Enter through Room1 North door, get hot water from kettle and add to tea, add milk from fridge, get snacks from NE cupboard, then sit at dining table



(a) GHWTV, SWTV, AHWTV, RN, HB-T, HB-E

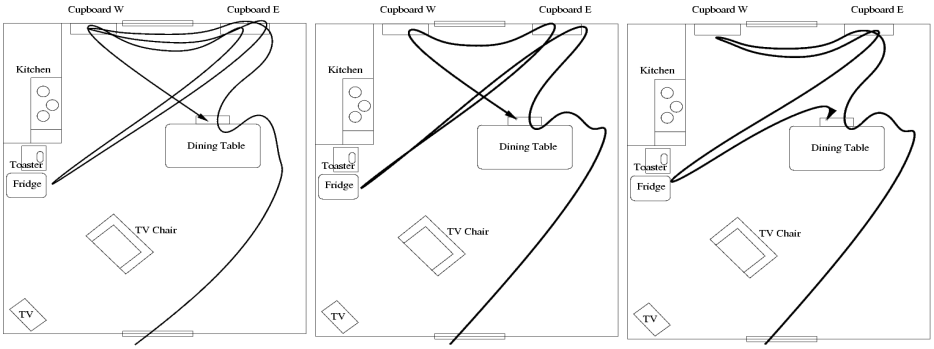


(b) CU, HB-P, HB-C, AT

Figure 2.11: Spatial paths of dataset A activities.

2.6.3 Dataset B (3 Activities)

The purpose of dataset B is to represent spatially similar activities captured in a mock smart home environment. Activities are approximately 90 seconds in length, with subjects entering from the S door of Room 1 and carrying out the specified activities in a consistent manner. Video is captured at 10 fps and processed by the video tracking system of Nguyen *et al.* (2002). The dataset comprises 3 single-person variants of having breakfast and contains a significant number of outliers. Figure 2.12 shows the spatial paths taken for the activities.



(a) HB1 (b) HB2 (c) HB3

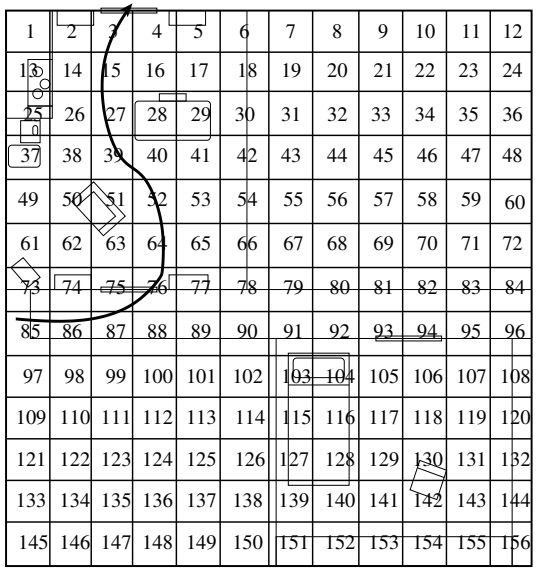
Figure 2.12: Spatial paths for dataset B activities.

2.6.4 Dataset C (12 activities)

The purpose of dataset C is to represent a diverse group of spatial activities captured in the same room within a mock smart home environment (reducing multi-camera tracking noise). Activities are approximately 60-90 seconds in length with subjects entering the smart home from the N or S door of Room 1. The dataset comprises 12 single-person activities which are similar to those carried out in datasets A and B. Video is captured at 10 fps and also processed by the tracking system of Nguyen *et al.* (2002).

2.6.5 Discretisation of Spatial Sequences

The tracking system used in this thesis outputs relative x, y coordinates of objects within the environment; however, approaches such as the discrete HMM and LCSS require one dimensional symbolised sequences. To obtain these sequences, the smart house environment is discretised into one square metre grids with each activity sequence, composed of x, y trajectories, being mapped to a sequence of unique integers u, where $u \in U$ and $U = 1, 2, 3, \ldots, 156$. This process can be seen in Fig. 2.13). As dataset B and



(a) Discretised smart house environment

$$[(0.5, 7.2)(1.5, 7.2)(2.3, 7.1)(2.8, 6.9)(3.5, 6.5)(3.6, 5.5)$$
$$(3.5, 4.5)(3.1, 3.8)(2.6, 3.6)(2.3, 2.6)(2.2, 1.4)(2.5, 0.5)] \Longrightarrow$$

(b) Two dimensional sequence

[85,86,87,75,76,64,52,40,39,27,15,3]

(c) One dimensional sequence

Figure 2.13: Mapping of smart house trajectories to symbolic forms.

C are localised to Room 1, the x, y trajectories are mapped into integers u, where $U = 1, 2, 3, \ldots, 72$ in accordance with Fig 2.14.

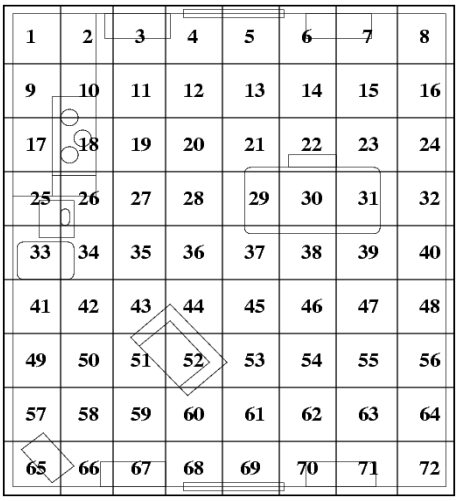


Figure 2.14: Discretisation Grid for Room 1 of the Mock Smart House Environment.

2.7 Summary

This chapter has presented a review of related work and background for research in this thesis. The chapter begins describing human activities and their characteristics, outlining the complexity associated with trying to robustly recognise human activities. A brief outline of smart homes is then provided to give context to the problem of spatial activity recognition in a smart home context. Following this, related work in human activity recognition is provided, focusing on template recognition approaches such as DTW, EDR and ERP, and the HMM, which is used in comparative trials. Next, an overview of bioinformatics sequence alignment approaches is given to provide a basis for which to discuss the novel approaches developed in this thesis. Existing biological paradigms that have been applied to computing and their rationale are then outlined in the following section. Finally, the smart home setup and data sets utilised in the subsequent chapters are detailed.

The following chapters will explore the application of a biological paradigm and bioinformatics-inspired sequence alignment approaches to the problem of spatial activity recognition in the presence of noise and activity variability.

Chapter 3

A Chemotactic Paradigm for Recognising Simple and Interwoven Spatial Activities

Dealing with noise incorporated during video tracking is challenging in spatial activity recognition. Factors such as shadows, reflections, lighting variation, and occlusions in such an environment have been shown to affect the accuracy of spatial coordinates generated by these systems (Nguyen *et al.*, 2003; Peursum *et al.*, 2003). The issue of accurate tracking is further exacerbated when multiple people are required to be tracked in the same smart home environment, with tracking systems typically using an individual's characteristics, such as calculated height, width, texture, colour of clothing, skin luminescence and/or velocity, to correlate positions with people.

Using the representation outlined in Aggarwal and Park (2004), activities are interpreted in this chapter at a macro level, where the focus is on determining movement patterns of individuals. As video data is captured in a smart home across multiple rooms, the tracking noise is typically large (even with smoothing) in relation to relative position of an individual and the size of the smart home. The magnitude of noise in this scenario is more significant in regards to activity discrimination, than if present in surveillance situations in large outdoor areas. It is important that spatial activity recognition approaches in smart homes can tolerate this variation without impacting discrimination.

In addition to the complexity of noise, human activities are routinely interwoven for efficiency, making recognition of activities more difficult. The switching between activities is difficult to model from a computational perspective, requiring state representation and a mechanism to switch between models. Recently, Kim *et al.* (2010) has provided a skip-chain conditional random field (CRF) and HMM approach that is capable of recognising simple concurrent and interwoven activities. A differential signature-based approach

is also provided that is used to data mine sensor data for concurrent and interwoven activities.

Biological systems have proven to be a useful basis for solving many real world problems as a consequence of their innate parallelism, robustness, adaptiveness, diversity and error tolerance (Paton, 1994). To develop a robust approach for dealing with tracking system noise and recognition of interwoven activities, the biological paradigm of bacterial chemotaxis is explored and a model formulated based on the robust and multi-cellular characteristics of the process. Bacterial chemotaxis is an appropriate paradigm for modelling as it allows bacteria to tolerate dynamic environments, improving survivability. This robustness characteristic is the key motivation for this work.

The cellular chemotactic model is derived in this chapter and evaluated in a smart home context. The template-based model is shown to exhibit high classification accuracy (99%), outperforming the discrete Hidden Markov Model (HMM) with a ten class activity set. High accuracy ($> 89\%$) is also maintained across small training sets and through incorporation of varying degrees of artificial noise in test sequences. Importantly, unlike other bottom-up spatial activity recognition models, the chemotactic model is demonstrated to be capable of recognising simple interwoven activities.

The following chapter is organised as follows. An overview of bacterial chemotaxis is covered in 3.1 as the inspiration for the activity recognition model. Section 3.2 describes the cellular chemotactic model and how it is applied to the activity recognition problem. The experimental methodology for investigating the robustness, recognition performance and ability to recognise interwoven activities is covered in section 3.3. Section 3.4 describes results from the investigation of the cellular chemotactic model and its application to activity recognition. The model is evaluated against a discrete HMM approach to determine classification accuracy with varying magnitudes of noise and interweaving of activities. Lastly, a summary is presented in section 3.5.

3.1 Bacterial Chemotaxis

Chemotaxis is a process that increases survival of motile bacteria such as *Escherichia coli* and *Salmonella typhimurium* by allowing the organisms to directionally swim in

response to chemical or other physical gradients (Adler, 1975). A chemotactic bacterium moves towards nutrients and increasing nutrient gradients, but is repelled from harmful environments and increasing gradients of harmful substances (Bourret and Stocks, 2002). Passino (2002) describes this respective behaviour as similar to a saltatory search.

In a uniform and static environment, chemotactic bacteria carry out a random walk by modifying flagella[1] motion, where the motion consists of alternating tumbles (changing direction) or running (going forward in a straight line)(MacNab and Koshland, 1972; Adler, 1975). The duration of tumbles and runs are exponentially distributed, with the mean duration of runs being approximately ten times longer than that of tumbles, allowing cells to "walk" (Berg, 1990). In the presence of an increasing favourable gradient, bacteria decrease the tumbling frequency and increase the run length allowing organisms to move toward an attractant source and thus feed Segall *et al.* (1986).

Motile bacteria interact with their environment through a process involving cell surface receptors that monitor environmental conditions. Binding of molecules to these cell receptors produces a signal that allows cells to respond to an input stimulus (Adler, 1975). Research on *Escherichia coli* has shown that bacteria can sense spatial gradients through temporal changes in attractant or repellent concentration at their receptors (MacNab and Koshland, 1972; Segall *et al.*, 1986). This eludes to the possibility of an intracellular short term memory, that allows cells to remember previous spatial concentrations for comparison to current levels (Segall *et al.*, 1986). The extracellular to intracellular mapping process, termed signal transduction, integrates extracellular signals (for example the binding of a molecule to a receptor) translating them into a series of intracellular structural or chemical changes. Through a series of enzyme catalysed reactions, cellular production levels or function are in turn modified (Bourret and Stocks, 2002). In the case of *Escherichia coli* and bacterial chemotaxis, the resulting change affects the "motor" of the flagella, increasing the duration of clockwise flagella rotation, resulting in a longer run movement towards a favourable environment (Berg, 1990; Passino, 2002).

It is well established that chemotaxis confers an evolutionary advantage to bacterial species that possess the characteristic, allowing them to survive and respond to changes in dynamic environments (MacNab and Koshland, 1972). From a systems perspective, bacterial chemotaxis is a robust process, showing temporal sensitivity to changes in nutrient concentrations even at different concentration levels. This sensitivity allows

[1]Flagella are whip-like appendages that provide locomotion similar to a propeller

motile bacteria to find and remain in nutrient rich environments improving survivability and fitness (Barkai and Leibler, 1997; Alon *et al.*, 1999). Even in areas with high concentrations of nutrients, motile bacteria will only remain in that vicinity for a duration before returning to a random walk and saltatory search Passino (2002). This behaviour allows chemotactic bacteria to continually search for areas with higher levels of nutrients (preventing fixation at local maxima), improving an organism's robustness to changing environments and thus survivability. This process also has parallels with the technique of simulated annealing (Kirkpatrick *et al.*, 1983) used in global approximation problems.

3.2 Cellular Chemotactic Model

An environmental and cellular abstraction of chemotaxis is used to derive the cellular chemotactic model, for addressing robustness issues with spatial activity recognition. The model formulation and parameters are outlined as follows. Activities are represented as cells in the model, and cells exist and move in an environment E, which is a two dimensional discretised space as per Fig. 2.13. A cell c has a type or activity label t, with a group of j cells (comprising the representative class set) having a representation of c_j^T. Concentrations of molecules (symbols) in the spatial environment are maintained using a histogram E_h with n bins, where n is the number of different symbols that need to be recognised and $n = |E_h|$. Each activity type is represented by a group of cells, where the cells model the movement of individual bacteria in response to environmental dynamics. Cells are composed of receptor types $\{R_i\}_{i=1}^n$ that match symbols from the environment. Symbols are denoted by u, where $u \in U$ and U is the set of all possible spatial symbols, for example $U = \{1, 2, 3, \ldots, 156\}$. Each receptor type has a specified number of receptors denoted by $|R_i|$. The total number of receptors of a cell is given by p according to (3.1):

$$p = \sum_i |R_i| \tag{3.1}$$

Cells have an x, y tuple parameter that determines the activity cell position within the environment E and are represented as $c(x, y, \{R_i\})$, where $i = 1, 2, \ldots, n$. These values are initially set to 1.0 and 0.0, respectively and represent the starting position of the cells in the environment. In the model, the attractant source s or the place where molecules are conceptually released is set to the origin of the environment, that is $s = (0.0, 0.0)$. This is graphically depicted in Fig. 3.1. Cellular running times for cells

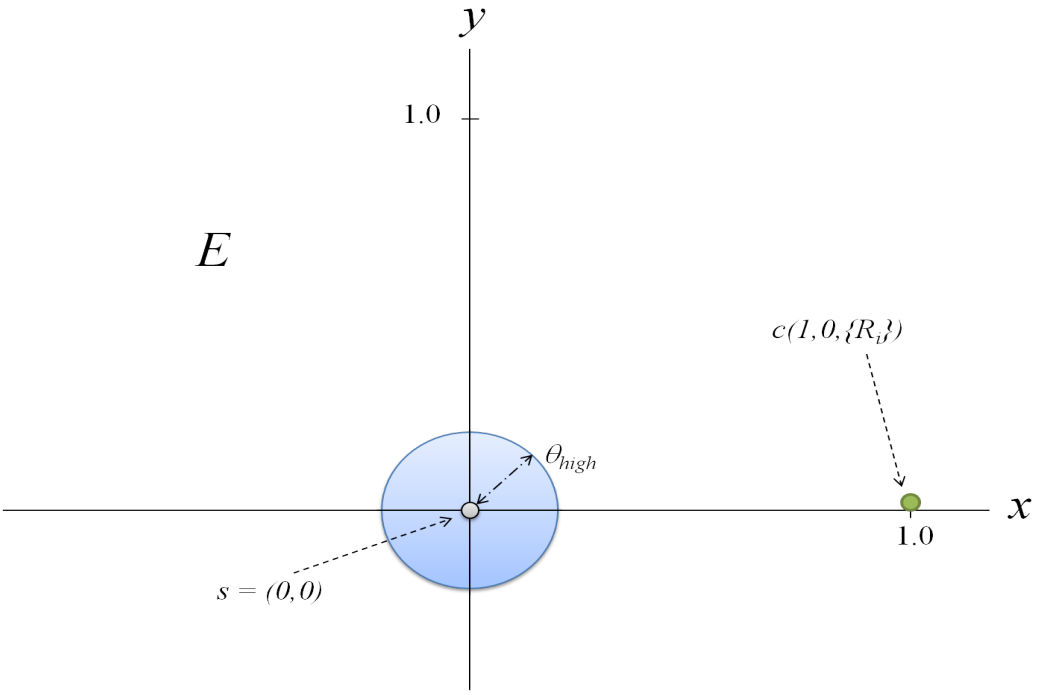


Figure 3.1: Cellular Chemotactic Environment E outlining the origin for cells c, the attractant origin s and the high concentration region governed by θ_{high}

with and without environmental gradients are represented by exponential distributions with means of μ_{LR} and μ_R, where $\mu_{LR} > \mu_R$. The exponential mean of the distribution for a long running movement is denoted by μ_{LR}, with μ_R being the exponential mean for the normal running motion. The amount of movement of a cell in response to a matching symbol is determined by the velocity v of the cell, according to (3.2). Velocity is normalised against the number of receptors p and μ_{LR}, to account for variation in inter and intraclass sequence length. The normalisation is important as it only allows cells with the closest match to the observed pattern to move nearest to the origin, thus reducing misclassification.

$$v = \frac{1}{p \times \mu_{LR}} \quad . \tag{3.2}$$

In normal biological settings the release of molecules into a fluid environment results in formation of a gradient that dissipates over time. This gradient is traversed by bacterial cells in order to locate an attractant source, thus increasing the fitness and

survivability of the organism. In the chemotactic model, iterative completion of a test sequence and the mapping of this to a symbol (molecule), results in the addition of $u \in U$ into the environment E_h at time t. The increase in environmental u, modelled as an increase in the environment histogram bin frequency of E_h, is detected by activity cells c with free receptors R_i of the same receptor type. Typically, chemotactic bacteria would then make a series of random moves with a bias toward the attractant source, re-evaluating environmental concentrations at each move. In this model, the area of highest concentration of molecules or the attractant origin is known, therefore, cells can easily determine the direction of travel and move toward the attractant if and only if receptor types match, free receptors are available and the cell is not in a region of high concentration. The process of symbol matching and non-matching of a cell c in an environment E with environmental gradient E_h is depicted in Figures 3.2,3.3 and 3.4.

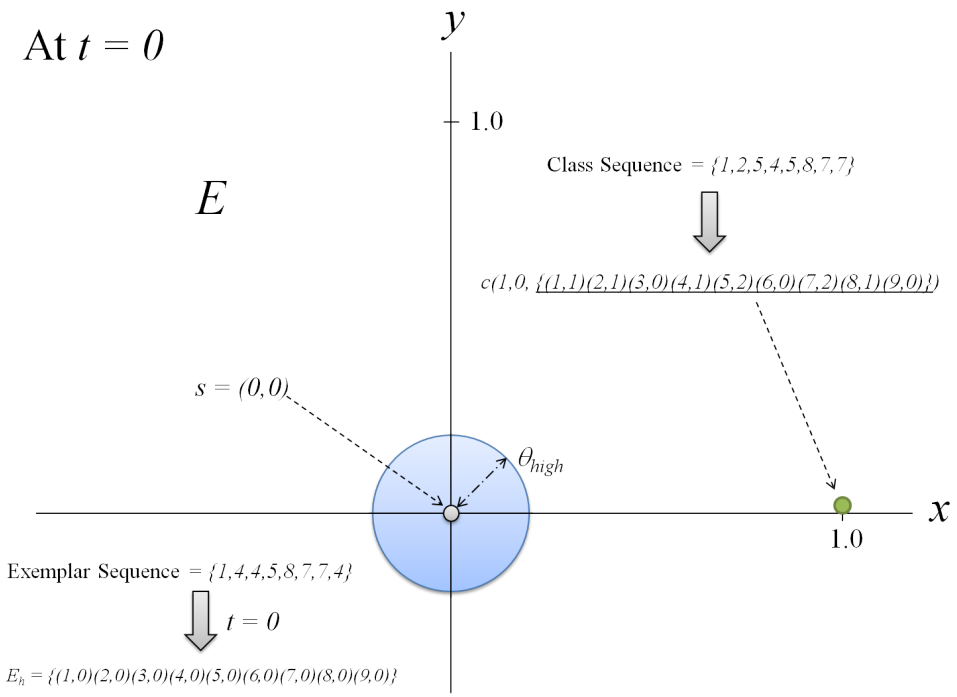


Figure 3.2: Graphical representation of the symbol matching process at $t = 0$ (initialisation)

Chemotactic cells in the model detect increasing environmental concentrations via "memory" associated with the irreversible binding of molecules to receptors. In the model the

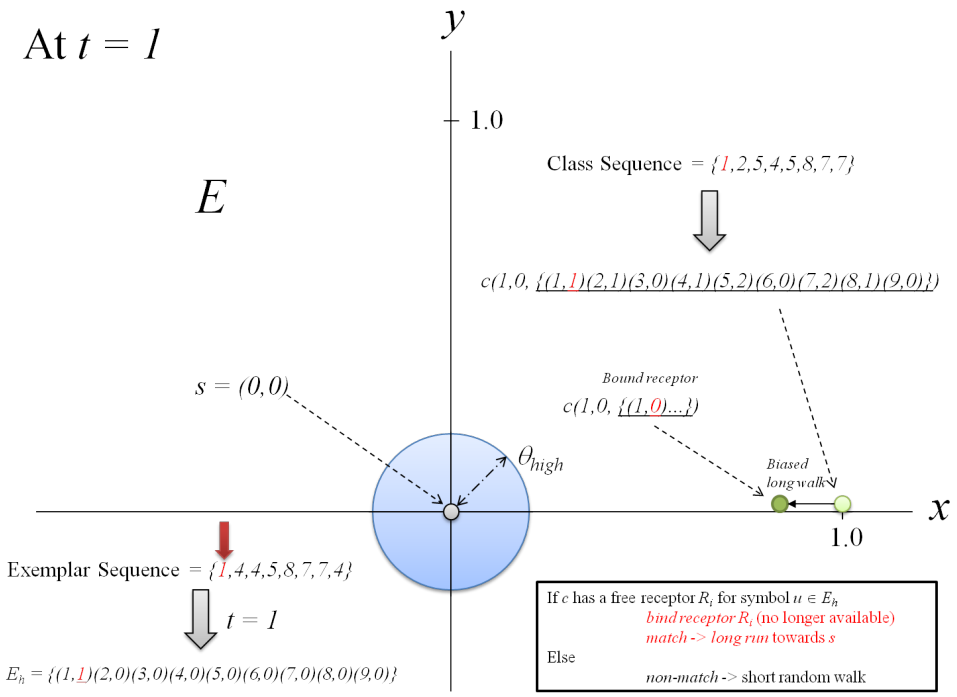


Figure 3.3: Graphical representation of the symbol matching process at $t = 1$ (match)

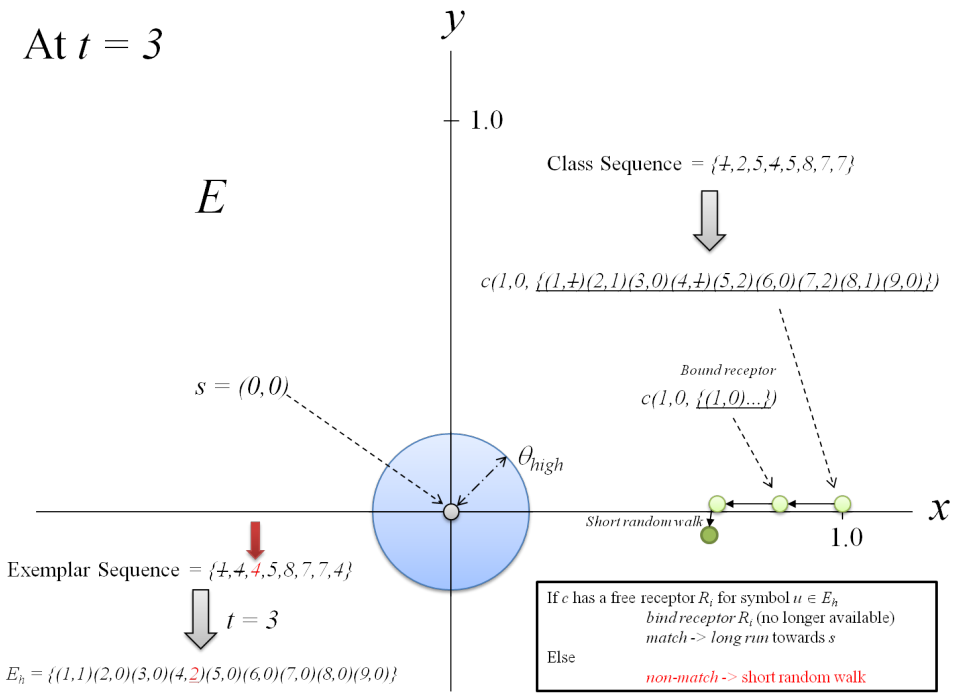


Figure 3.4: Graphical representation of the symbol matching process at $t = 3$ (non-match)

cellular memory is represented with a histogram approach with selected histogram bins representing the receptor types. Each histogram bin or receptor type has a corresponding fixed maximum bin frequency. The maximum bin frequency describes the number of receptors $|R_i|$ of a given receptor type R_i that a cell possesses. When a molecule u is released into the environment, the environmental concentration of that chemical increases. Cells that have a matching receptor type for the molecule then check if any of the particular receptors are free. If so, the molecule binds and the cells behaviour is modified by changing direction toward the attractant origin and increasing the length of the running movement governed by μ_R to μ_{LR} as shown in 3.3. If a cell does not have a receptor for that particular molecule or the cell does have a corresponding receptor type but no free receptors, then a random walk is performed over a distance obtained from the exponential distribution with mean μ_R. When cells move close to the attractant source and the euclidean distance d between the cell and origin is less than or equal to the high concentration threshold θ_{high}, the cells perform random walks irrespective of increasing environmental concentrations. If the cells move outside the high concentration area where $d > \theta_{high}$, then the cells return to a normal random walk behaviour. These behaviours are shown more clearly through an example in Figures 3.5 and 3.6. Motile bacteria use this corresponding change in behaviour at high environmental concentrations to prevent being restricted to local regions of high attractant concentrations (Barkai and Leibler, 1997; Alon *et al.*, 1999). In the model, this behavioural characteristic is used as a tolerance mechanism for sequence expansion.

Figure 3.7 illustrates the behaviour of similar and dissimilar activity cells in response to an observed spatial sequence (or environment E). From Fig. 3.7 it is clear that activity cells with similar patterns to the test sequence exhibit more straight line movement toward the attractant source compared to dissimilar activity cells. Therefore, cells c with higher degrees of similarity to test sequences will end up closer to the attractant source s. From a qualitative perspective, areas of sequence similarity may also be heuristically identified through visualisation of stretches of straight line movement, that is before, after or between regions of random walking.

After all symbols, which are mapped from the trajectories of an activity sequence, are iteratively transformed into cells c^{β}, comprising β classes with m cells per class, the cell ϕ in Z is found, where Z is the set of all activity cells, and ϕ has the minimum Euclidean

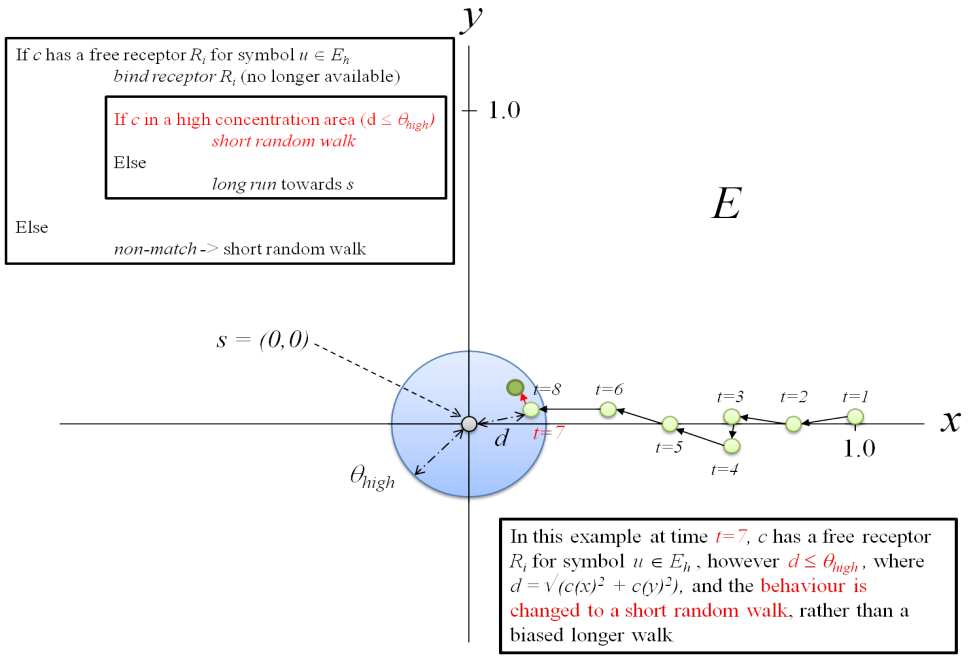


Figure 3.5: Example of a short random walk behaviour when cells c are close to an attractant ($d \leq \theta_{high}$)

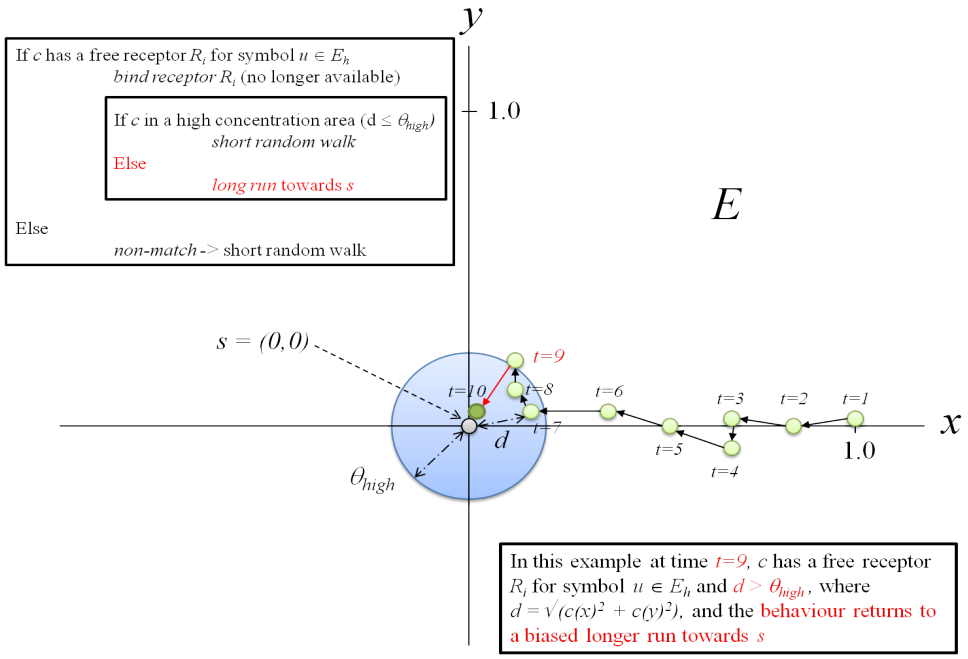


Figure 3.6: Example of a return to normal behaviour when cells c move away from an attractant ($d > \theta_{high}$)

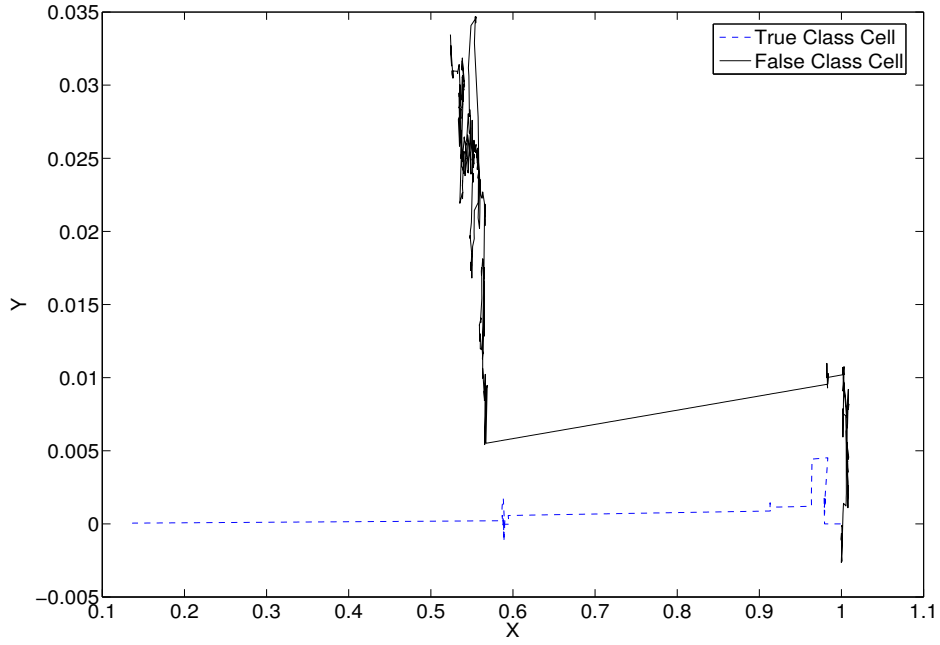


Figure 3.7: Chemotactic Cell Movements of true and false class cells. Movement is in the direction of the origin.

distance to the attractant source δ of E according to (3.3).

$$\phi = \underset{g \in Z}{\operatorname{argmin}}\, d(g, \delta) \quad . \tag{3.3}$$

The minimum distance cell ϕ is then used in the classification decision.

3.3 Methodology

For the evaluation, a ten single-person activity dataset (dataset A) was utilised comprising the following activities: *get home-watch TV*, *have a snack-watch TV*, *at home-watch TV*, *read newspaper*, *have breakfast-toast*, *have breakfast-eggs*, *clean up*, *have breakfast-porridge*, *have breakfast-cereal* and *afternoon tea*, with twenty sequences captured per activity (further information on the dataset and its corresponding actions can be found

in 2.6). To generate symbolic sequence representations, the smart house space was discretised into one metre grids and activity sequences, composed of x, y trajectories, were then mapped to a sequence of unique integers u, where $u \in U$ and $U = 1, 2, 3, \ldots, 156$ as per Section 2.6.

Following mapping of the activity sequences, sequences were randomly separated into training and testing sets for each of the $\beta = 10$ activity classes. Each test sequence was then allocated a separate environment E, with environmental histogram E_h of size $|U|$, with bins initialised to zero. For each environment, the training sequences i ($i \leq m$) of each activity class j ($j \leq \beta$) were then transformed into cells c_i^j and initialised according to algorithm 1: Cell velocity was then derived for each cell c_i^j using (3.2). To quantify

Algorithm 1: Creation and initialisation of cells c_i^j from training sequences

for $i \leftarrow 1$ **to m do**
  for $j \leftarrow 1$ **to** β **do**
    /* *Initialise cell location to (1,0)* */ ;
    $c_i^j \leftarrow (1, 0, \{\})$;
    /* *Convert discretised training sequence i of activity class j to a receptor vector* $\mathbf{R}$ *of size* $|\mathbf{R}| = |U|$ */ ;
    for $k \leftarrow 1$ **to** $|U|$ **do**
      /* *a. Initialise vector at index k* */ ;
      $R_k \leftarrow 0$;
      /* *b. For each symbol k in sequence i, j increment the number of receptors at R_k* */ ;
      $R_k = R_k + 1, \forall k \in seq(i, j)$;
    /* *Assign receptor vector* $\mathbf{R}$ *to initialised cell i, j* */ ;
    $c_i^j(1, 0, \mathbf{R})$;

the recognition performance of the approach with the testing sets, cross-validation was performed, allowing generation of random training and testing sets to provide a more realistic interpretation of discriminatory performance.

To benchmark the effectiveness of the chemotactic approach in relation to other models the chemotatic model was evaluated against a discrete HMM with consequent HMMs built for each activity class (using the same training sequences). HMM models were trained using Baum-Welch parameter estimation, with the number of iterations of the algorithm controlled by the convergence of the ratio of the average of the log-likelihoods

between the current and previous iterations (< 0.001)). The ability of the chemotactic model to function adequately with few training sequences or templates was also evaluated, with results compared to the HMM. The intention of this was to determine the viability of using limited training templates in order to accurately recognise spatial activities. Models such as the HMM typically require larger training sets in order to produce good discriminative models.

Following from this, the robustness of the model was measured through introduction of varying magnitudes of Gaussian noise across the testing sequences, prior to the symbol mapping process. Introduction of artificial noise into the sequences allows one to make an empirical comparison of robustness in relation to the HMM. As chemotaxis is inherently robust to changing environmental conditions, it is believed the model would show similar robust characteristics to noise.

The final evaluation of the model was in a simple interwoven activity context, where two activities were used to contrast the ability of this novel approach to deal with activity interweaving. Recognition of interwoven activities has been seldom addressed in activity recognition research up to this point due to the difficulty of modelling activities subject to interruption. Recognition of interwoven activities is necessary in any real life activity recognition system as humans typically interweave activities to achieve temporal efficiencies. This experiment empirically demonstrates how the chemotactic model can transition to a random short walking state during interruptions in activities and then resumes with biased longer runs towards an attractant origin, when activities recommence.

3.4 Experimental Results

3.4.1 Parameter Selection

The chemotactic model has three parameters that affect recognition performance that require empirical optimisation: μ_R, μ_{LR} (govern the exponential distributions controlling normal running and long running cell movements) and θ_{high} (controls which running motion is taken in areas of high concentration). Given the considerations outlined in

Section 3.2, an empirical evaluation was performed with ten training sequences per class to determine the set of optimal model parameters (optimal in relation to recognition performance).

In determination of optimal values of μ_R and μ_{LR}, the effect of θ_{high} was minimised by setting $\theta_{high} = 0$. Different values of μ_R and μ_{LR}, where $\mu_R \leq \mu_{LR}$, were then specified for the models. Accuracy statistics were in turn obtained through cross-validation analysis. It was found that the ratio of optimal values of μ_R to μ_{LR} were consistent across different values of the exponential means, with μ_R typically one third of μ_{LR}. In Berg (1990), the bacterial ratio was found to be approximately a tenth. The consistent ratio can be explained in part by the cell velocity in (3.2) taking into account μ_{LR}; therefore, any increase in μ_{LR} inversely affects cell velocity, slowing the movement of the cell and approximating a smaller random biased walk. When values of μ_R approach μ_{LR} a decrease in accuracy is noted, which can be attributed to a lack of disparity between the matching and non-matching states of the cell. Using the acquired optimal ratio, the following parameters were included in further experimentation: $\mu_R = 0.5$ and $\mu_{LR} = 1.5$.

With the optimal movement parameters, values of θ_{high} between 0.01 and 0.10 were evaluated to ascertain an optimal range. The optimal range was found to exist between θ_{high} values of $0.03 - 0.05$; however, the difference in accuracy between the values was found to be less than 1%. This small difference in accuracy with different values of θ_{high} is likely due to the approach obtaining near 100% correct classification with the given dataset. In an activity recognition context, the θ_{high} threshold is used to change movement behaviour, keeping cells that are already close to the origin (equating to a high probability of a match) in the vicinity, irrespective of further elements in the testing sequence. Therefore, spatial sequences exhibiting expansion due to an activity occurring over a longer duration can benefit from the inclusion of the θ_{high} parameter and still be recognised as the same activity with a shorter duration.

The discrete HMM was included in the study to provide a benchmark comparison for the chemotactic approach in relation to an existing activity recognition technique. The discriminatory and runtime performance of the HMM is correlated to the number of hidden states used in the generated model. To address this issue, the HMM was empirically evaluated with hidden states $= 5, 7, 12, 15$ against the data set. The optimal classification accuracy was obtained where $M = 5$ as seen in Table. 3.1.

M	**Accuracy(%)**	**Stdev**
5	88.93	3.28
7	86.87	4.08
12	88.50	4.18
15	87.10	3.82

Table 3.1: Classification accuracy of discrete HMM models with $M = 5, 7, 12$ and 15 hidden states.

3.4.2 Recognition Performance

Using the following derived optimal parameters, $\mu_R = 0.5$, $\mu_{LR} = 1.5$, $\theta_{high} = 0.03$ and $M = 5$, chemotactic and HMM models were generated from ten training sequences according to the methodology described in Section 3.3. Ten test sequences per activity type were then evaluated in regards to recognition accuracy with the results shown in Table. 3.2. As evident from Table 3.2 the chemotactic model showed a significant $\approx 11\%$

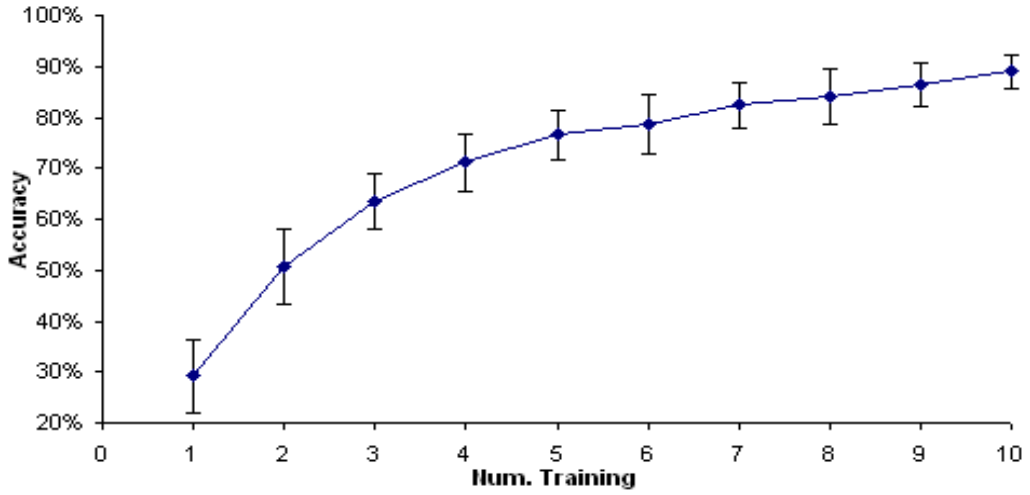
Technique	**Accuracy (%)**	**Stdev**
Chemotactic Model	99.89	0.36
HMM	88.93	3.28

Table 3.2: Classification accuracy for the Chemotactic and HMM models.

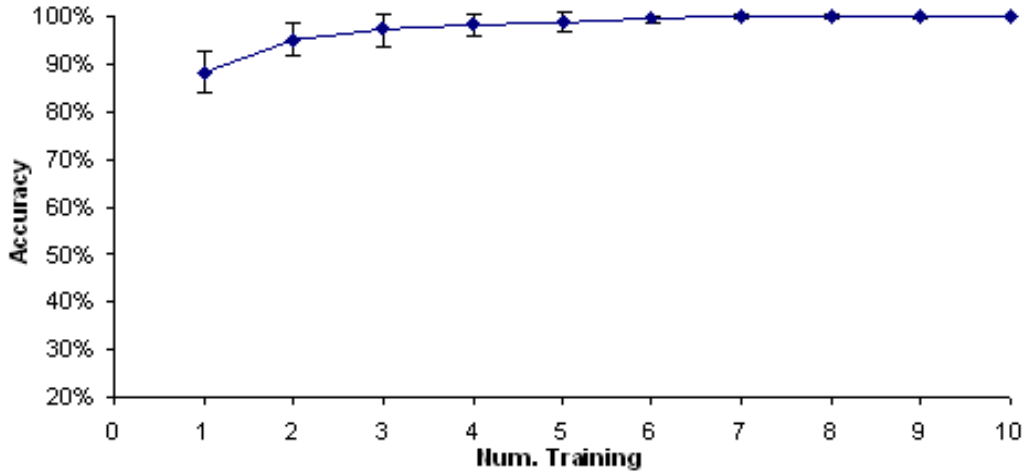
improvement in classification accuracy over the HMM with the ten activity dataset. The observed higher recognition performance of our model is the result of the chemotactic process better accounting for tracking noise through its random walk motions in the absence of sequence correspondence. Additionally, even though sequential consistency is not enforced in the chemotactic model, the model still exhibits high discriminative properties with sequences of similar length.

3.4.3 The Effect of Training Size on Recognition Performance

To analyse the effect of increasing numbers of training sequences with respect to classification performance, chemotactic and HMM models were trained with one to ten sequences of each activity from the ten activity dataset. Figure 3.8. demonstrates the effect of increasing training sequence numbers in relation to accuracy.



(a) HMM



(b) Chemotactic

Figure 3.8: Number of training sequences versus accuracy for the HMM and chemotactic models. The bars represent one standard deviation from the mean.

It is apparent from Fig. 3.8 that the chemotactic model is able to recognise activities with significantly higher accuracy and less variation than the HMM, especially with smaller numbers of training sequences. The chemotactic models good performance is attributed to the noise abating characteristics of the chemotactic paradigm and the generalisation capability of the underlying histogram approach. The high recognition performance with

small training set sizes further demonstrates that the proposed chemotactic approach does have high discriminatory properties. Furthermore, this characteristic is particularly useful for recognising activities with limited sequence data.

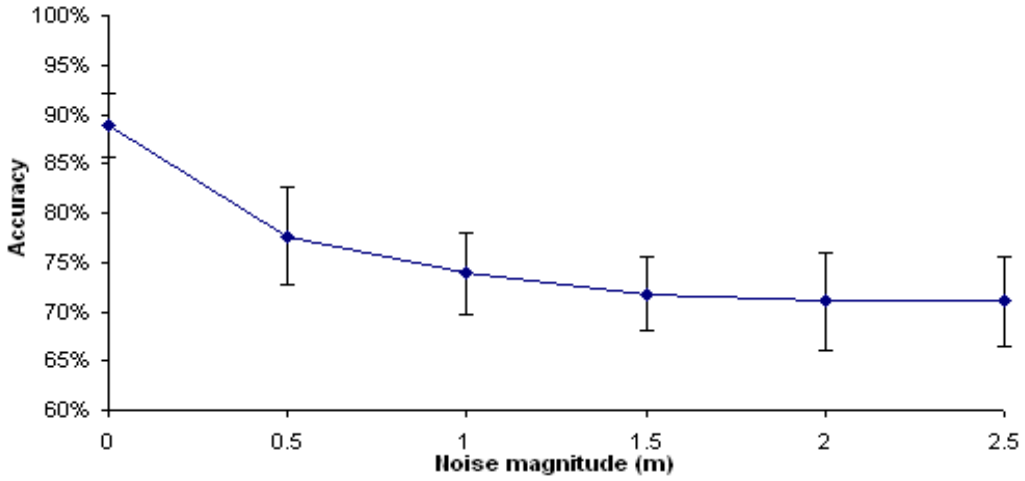
The lower classification accuracy obtained by the HMM with smaller numbers of training sequences was expected as the discrete HMM is unable to accurately derive the probabilities of the state transitions with smaller training sets, thus resulting in a poor recognition performance.

3.4.4 Noise Tolerance

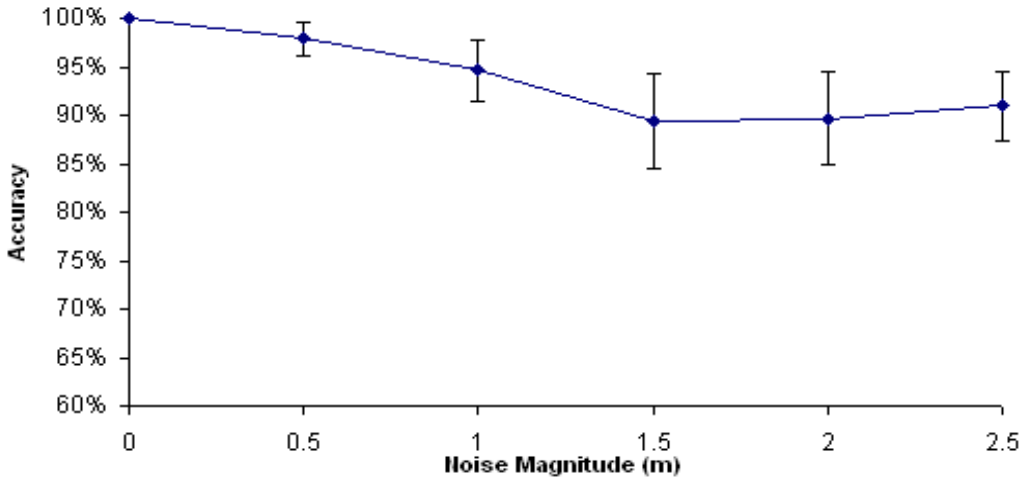
To verify the robustness claims of the chemotactic model in relation to the HMM, Gaussian noise was artificially incorporated with varying degrees of magnitude into the two dimensional sequence data, prior to the symbolic mapping process. Chemotactic and HMM models were then generated with ten training sequences of the original sequence data and analysed for classification accuracy using the ten "noisy" testing sequences per activity. The experimental results with noise magnitudes of between 0 to 2.5 metres are graphically represented in Fig. 3.9. Analysis of the results show that the chemotactic model is more resilient to noise than the HMM, as indicated by the lesser decrease in accuracy (10% versus 17%) with increased magnitudes of noise. Furthermore, the chemotactic model still maintained 90% classification accuracy with the largest degree of noise. This contrasts the HMM which only obtained 71%. Both models did however produce similar amounts of variation in the experiments as seen by the respective standard deviations. The observed robustness of the chemotactic model can be explained by the approach biasing longer duration random walks of activity cells (in the direction of an attractant source) when patterns match, whilst conducting shorter unbiased random walks with sequence expansion and dissimilarity. The higher accuracy of the chemotactic model with artificially introduced noise acknowledges the robustness of the approach.

3.4.5 Recognising Simple Interwoven Activities

To validate the ability of the chemotactic model to recognise simple interwoven activities an alternate data set was constructed that consisted of two classes of activities as shown



(a) HMM



(b) Chemotactic

Figure 3.9: Noise magnitude versus classification accuracy for the HMM and chemotactic models. The bars represent one standard deviation from the mean.

in Fig. 3.10.

The first activity is *have a snack-watch TV*, with the second activity *cooking-eating* starting at the dining table, involving some cooking and then returning to the dining table for eating. Chemotactic cells were trained with both classes of activities and

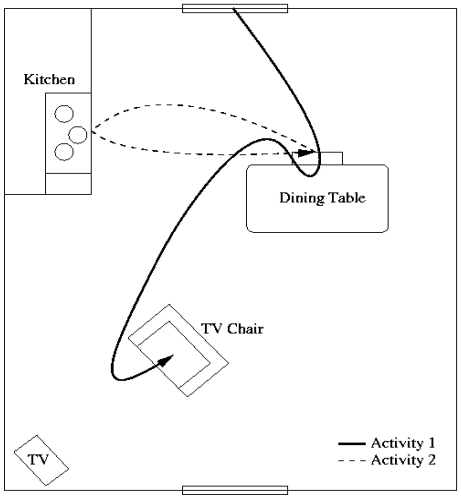


Figure 3.10: Snack-Watch TV and Cooking-Eating spatial paths.

evaluated with an interwoven sequence that involved entering from the north door of room 1, having a snack, cooking, returning to the dining table to eat and then proceeding to watch tv. The interwoven sequence thus incorporates the beginning section of activity 1, transfers to activity 2 and then returns to complete activity 1. The overall length of the interwoven sequence is the sum of the lengths of activity 1 and 2. Figure 3.11. shows the resulting cellular movements of both classes in response to the interwoven sequence.

Evaluation of Fig. 3.11a, showed a long random walk at approximately $x = 0.85$, indicating interruption of activity 1. At the same time ($x \approx 1.0$), the other activity cell in Fig. 3.11b was observed to change from a random walk to straight line movement toward the attractant origin (0,0) indicating similarity to the test sequence. The cell in Fig. 3.11b then reverted back to a random walk at $x = 0.05$, with the opposite cell in Fig. 3.11a returning to straight line movement toward the attractant origin. As demonstrated in Figures 3.11a and 3.11b, the final positions of both the SWTV and Cooking cells were close to the attractant origin showing that both activities had been recognised.

To verify that the HMM is unable to recognise such activities, separate HMMs were trained with the same two classes of activities and then used the above interwoven sequence for testing. The resulting log likelihoods of the models with the observed interwoven sequence were found to be negative infinity. This occurs as the models do

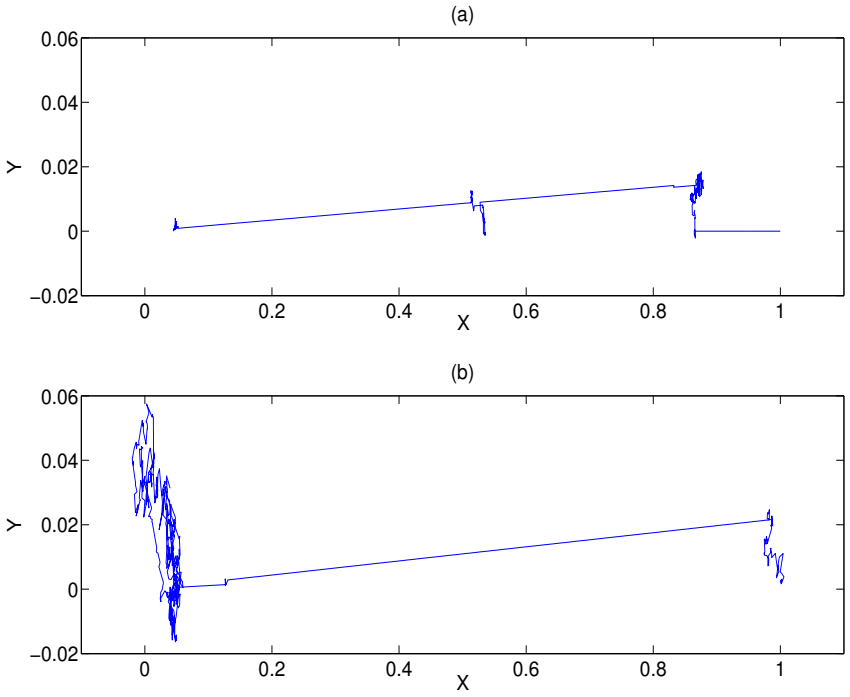


Figure 3.11: Interwoven cell movements (a) Snack-Watch TV Cell (SWTV) (b) Cooking Cell.

not observe particular states, found in the other class during training. Therefore, when calculating the sequence probability with a scaled forward procedure (Rabiner, 1989), some of the derived symbol probabilities for observed symbols are zero or very small, resulting in an overall low probability for the sequence.

The chemotactic approach has therefore been demonstrated to recognise simple interwoven activities. The exhibited tolerance to interweaving is the result of the chemotactic cells performing smaller random walks in the absence of sequence similarity. Therefore, if an activity is in progress and a corresponding cell is moving in the direction of the attractant source (due to patterns matching), interruption of that activity will result in the cell performing a random walk. The random walk allows cells to remain in the vicinity of where the activity was disrupted, until the activity is resumed. If an activity is disrupted for a long period of time, it is possible that cells may *wander* away from the area of disruption, decreasing overall recognition performance.

3.5 Summary

This chapter has presented a novel chemotactic model for spatial activity recognition. The chemotactic paradigm that forms the basis of this model provides robustness, enabling tolerance to video tracking noise. This robustness is achieved by biasing longer duration random walks of activity cells in the direction of an attractant source, during sub-sequence matching and conducting shorter unbiased random walks with sequence dissimilarity. A discussion was provided on optimisation of the chemotactic model parameters in relation to empirical evidence. Using optimal parameters, the classification accuracy of the developed model with a ten activity dataset was determined, in which case the chemotactic model demonstrated higher classification accuracy (99%) than the discrete HMM (89%). Evaluation of the classification performance over different training set sizes showed that the chemotactic model also exhibited high discriminative properties, even when trained with few sequences. Through introduction of differing magnitudes of artificial noise into testing sequences, the robustness of the chemotactic model was demonstrated in relation to the HMM. These results suggest that the chemotactic paradigm is well suited to computational problems with intrinsic noise, such as the spatial activity recognition problem. Lastly in this chapter, the ability of the chemotactic model to recognise simple interwoven activities was verified.

Chapter 4

Improving robustness of spatial activity recognition using principles from bioinformatics

In Chapter 3, a novel cellular chemotactic model is described that is robust to noise associated with spatial activity sequences in a smart home environment and is also capable of recognising simple interwoven activities. This chapter further explores the challenge of recognising spatial activity sequences containing noise from video-based tracking systems using concepts from the field of bioinformatics for inspiration.

Multi-camera and other video-based tracking systems, such as the ones described in Section 2.6 introduce noise and artifacts into spatial sequences as a result of difficulties with tracking individuals in smart home environments. Tracking noise is seen as a localised variance, or as sequence compression or expansion in spatial cases. In biological sequences, localised substitution, compression and expansion of sequence elements are naturally occurring phenomena resulting from evolution. Consequently, bioinformatics techniques have been optimised to identify biological sequences exhibiting evolutionary variation. The proposed approaches addressed in this chapter are based on bioinformatics sequence alignment techniques (see Section 2.4) that are capable of quantifying sequence variability.

Recently, sequence alignment techniques have also been applied in other pattern recognition approaches such as matching moving object trajectories from video data (Vlachos *et al.*, 2002b,a; Chen *et al.*, 2004; Chen and Ng, 2004; Chen *et al.*, 2005). The approaches are capable of robust recognition of spatial activities, yet are sensitive to noise from tracking systems, have decreased discrimination abilities due to symbolic space transformations, are susceptible to time variance in activities, and/or are sensitive to activity segmentation limitations.

The following chapter is organised as follows. Nomenclature used throughout this chapter and the remainder of this thesis is provided in Section 4.1. In Section 4.2 and Section 4.3 the bioinformatics-inspired Longest Common Subsequence Distance (LCSD) and Global Edit Distance (GED) techniques are derived to address noise and temporal variance issues. These algorithms are then evaluated in Section 4.4 in relation to recognition performance and training set sizes. An evaluation is also conducted against LCSS variants, DTW and a discrete HMM. Lastly, a summary is presented in section 4.5.

4.1 Nomenclature

The following notation is used throughout this and other chapters. A symbol a_i or b_j represents a trajectory tuple (x, y), where x and y denote the position within a two dimensional tracking space. The sequences $\mathbf{a}$ and $\mathbf{b}$ with lengths $|\mathbf{a}|$ and $|\mathbf{b}|$ are composed of symbols organised in time sequential order, where i and j ($1 \leq i \leq |\mathbf{a}|$ and $1 \leq j \leq |\mathbf{b}|$) determine the position within that corresponding sequence. Thus, a sequence of symbols $\mathbf{a} = [a_1, a_2, \ldots, |\mathbf{a}|]$ can also be represented as a sequence of tuples $\mathbf{a} = [(x_1, y_1), (x_2, y_2), \ldots, (x_{|\mathbf{a}|}, y_{|\mathbf{a}|}])$ and in combination as in $\mathbf{a} = [(a_1^x, a_1^y), (a_2^x, a_2^y), \ldots, (a_{|\mathbf{a}|}^x, a_{|\mathbf{a}|}^y)]$. To simplify the notation the symbolic representation will be predominately used throughout this chapter.

4.2 Longest Common Subsequence Distance (LCSD)

The proposed Longest Common Subsequence Distance (LCSD) algorithm is a distance variant of the LCSS similarity approach (2.4.1). Like its name suggests LCSS finds the longest common subsequence from two sequences, where subsequences need not be adjacent, and the similarity score is given by the LCSS length. LCSS achieves robust recognition in the presence of noise by ignoring all regions of dissimilarity and maximising element matching between sequences. In sequence alignment applications, LCSS and variants are used to identify similarity between biological sequences that are distantly related from an evolutionary perspective and have large amounts of variation (indels). In activity recognition, it is important to accurately recognise disparate sequences through the intrinsic noise incorporated during video tracking. To achieve high levels of discrim-

ination between similar sequences, there is a requirement to account for dissimilarity. This is important as different sequences can have similar matching regions, but vastly different regions of dissimilarity. For example, if one measures the LCSS between a sequence $\mathbf{a} = [1234]$ and two observed sequences $\mathbf{b} = [1254]$ and $\mathbf{c} = [12744]$, both $\mathbf{b}$ and $\mathbf{c}$ will have the same LCSS score of three, indicating that both are equally similar to $\mathbf{a}$. However, by visual inspection, one would say that $\mathbf{b}$ is more similar to $\mathbf{a}$. To determine whether $\mathbf{b}$ or $\mathbf{c}$ is more similar to $\mathbf{a}$ requires quantification of the sequence dissimilarity. In this chapter the noise abating aspects of LCSS are incorporated and a new distance-based alignment technique is proposed. A similar LCSS-based approach has been proposed by Vlachos *et al.* (2002a), in which the authors incorporated a constant δ to restrict how far points can match in time and a matching threshold ϵ that specified how close trajectories must be in order to match. The resulting similarity value from the LCSS variant was then normalised using the minimum of the sequence lengths. The main disadvantage of the LCSS approach of Vlachos *et al.* (2002a) is that it completely ignores all areas of dissimilarity, which in turn may be important for discriminating between similar activities. Normalisation of this approach does incorporate an aspect of dissimilarity; however a more accurate means of dissimilarity quantification is required. The LSCD approach contrasts LCSS as it is distance-based and furthermore does not require an extra normalisation step. The distance-based approach is preferred over a similarity one to quantify the dissimilarity between regions of subsequence similarity. This is necessary as we assume trajectory patterns of the same activity to be similar and thus do not wish to match sequences which significantly divert from the norm.

LCSD employs a dynamic programming matrix C of size $|\mathbf{a}|+1 \times |\mathbf{b}|+1$, in conjunction with a matching threshold θ. The matching threshold θ controls the degree of similarity required for individual symbols to match, but inclusion of a matching threshold in LCSD violates the triangular inequality metric characteristic, as per the EDR approach of Chen *et al.* (2005). The use of θ in LCSD manifests as a spatial envelope of size θ surrounding the template sequence. If θ is large the envelope surrounding a template sequence is large and thus LCSD over generalises mismatching sequences with similar spatial profiles. If θ is small then the resulting spatial envelope is small resulting in sequences of the same activity class being not matched.

Selection of appropriate values of θ for sequence alignment based approaches are dependent on the spatial and temporal variability of exemplar sequences. As discussed in Section 2.1, the ADL routines and activity duration associated with individuals are

consistent for normal levels of well-being. Thus it is expected that activities captured from the same individual are normally distributed in relation to spatial and temporal variability. This was validated in early work through analysis of the spatial sequence distributions from Section 2.6.

A global threshold θ is favoured in LCSD and other sequence alignment approaches, and over local thresholds for each activity class, due to the ease which a global threshold can be empirically derived through sampling and consequently evaluating training data sets, to obtain a balanced precision and recall. In comparison to global thresholds, local thresholds are computationally expensive to calculate for nearest neighbour classification as the solution requires solving the following multi-factorial optimisation problem: for n activities and for each local threshold θ_k, where $k = 1, 2, \ldots, n$, find values of θ_k such that the precision and recall statistics 5.2) for all n are maximised.

The LCSD formulation uses an Euclidean distance function $d(a_i, b_j)$ (4.1) for quantifying symbol distances and a stepwise distance function $dist_{LCSD}(a_i, b_j)$ for DP initialisation.

$$d(a_i, b_j) = \sqrt{(a_i^x - b_j^x)^2 + (a_i^y - b_j^y)} \tag{4.1}$$

To derive the LCSD one initialises C using (4.2) and calculates the remainder of the matrix using (4.3), for values of $i = 1, \ldots, |\mathbf{a}|$ and $j = 1, \ldots, |\mathbf{b}|$.

$$C(i,j) = \begin{cases} dist_{LCSD}(a_i, b_j) & i = j = 0 \\ C(i, j-1) + dist_{LCSD}(a_i, b_j) & i = 0, j > 0 \\ C(i-1, j) + dist_{LCSD}(a_i, b_j) & i > 0, j = 0 \end{cases} \tag{4.2}$$

$$\text{where, } dist_{LCSD}(a_i, b_j) = \begin{cases} 0 & d(a_i, b_j) < \theta \\ d(a_i, b_j) & d(a_i, b_j) \geq \theta \end{cases}$$

$$C(i,j) = \begin{cases} C(i-1, j-1) & i, j > 0 \text{ and } d(a_i, b_j) < \theta \\ \min\{C(i, j-1), C(i-1, j)\} + d(a_i, b_j) & i, j > 0 \text{ and } d(a_i, b_j) \geq \theta \end{cases} \tag{4.3}$$

At each $C(i, j)$, one first determines whether the corresponding elements a_i and b_j match by calculating the Euclidean distance between the points and if the distance is less than the threshold θ. If so, then a match occurs and the current optimal distance value $C(i, j)$, is set to a match $C(i-1, j-1)$. If the Euclidean distance is equal to or exceeds

θ, $a_i \neq b_j$ and the non-matching elements are penalised through the addition of the Euclidean distance between the points to the minimum of $C(i-1,j)$ and $C(i,j-1)$. By including the Euclidean distance of non-matching elements we are able to accurately quantify the regions of dissimilarity between subsequences. The pseudocode for the LCSD algorithm is provided in Algorithm 2 for further clarity.

Algorithm 2: Longest Common Subsequence Distance (LCSD)

```
Data: θ, a, b
/* Initialise DP matrix C */ ;
for g ← 0 to |a| do
    for h ← 0 to |b| do
        C(g, h) ⟵ 0 ;
/* Initialise first row and column of DP matrix C */ ;
for i ← 1 to |a| do
    C(i, j) ⟵ C(i − 1, j) + dist_LCSD(a_i, b_j) ;
for j ← 1 to |b| do
    C(i, j) ⟵ C(i, j − 1) + dist_LCSD(a_i, b_j) ;
/* Perform LCSD calculation on remainder of rows and columns in DP matrix */ ;
for i ← 1 to |a| do
    for j ← 1 to |b| do
        /* Sequence elements match */ ;
        if dist_LCSD(a_i, b_j) < θ then
            C(i, j) ⟵ C(i − 1, j − 1) ;
        /* Not a match so find the next closest alignment */ ;
        else if C(i − 1, j) < C(i, j − 1) then
            C(i, j) ⟵ C(i − 1, j) + dist_LCSD(a_i, b_j) ;
        else
            C(i, j) ⟵ C(i, j − 1) + dist_LCSD(a_i, b_j) ;
```

To calculate the LCSD of one-dimensional spatial sequences, the matching criteria are replaced by the boolean expression, $a_i = b_j$, for a match previously determined by $d(a_i, b_j) < \theta$ and $a_i \neq b_j$ for non-matching cases previously represented by $d(a_i, b_j) \geq \theta$. The non-matching penalty, $d(a_i, b_j)$ is also replaced by a positive constant, which is set to one. The significance of the constant being one is that the resulting LCSD is in fact the number of symbols that are different between the two sequences.

4.3 Global Edit Distance (GED)

The Global Edit Distance (GED) approach that is outlined here is derived from the Edit Distance sequence alignment approach in bioinformatics. As such, GED in a symbolic context denotes the number of substitution or indel operations that are required to convert one sequence into another. This is achieved through minimising the alignment score consisting of the number of mismatches and indels and their corresponding penalties, over the sequence lengths $|\mathbf{a}|$ and $|\mathbf{b}|$ (Waterman, 1995). The GED algorithm proposed in this chapter differs from both LCSS and LCSD approaches as it allows indels or gaps in the sequences in conjunction with mismatches, and is distinguished from ERP (Chen *et al.*, 2004) due to the inclusion of a matching threshold in the distance calculation. Whether a match, mismatch or indel (corresponds to a deviation in a spatial path) occurs in the alignment is in turn dependent on the weights associated with each operation. As this application is dealing with two dimensional spatial sequences, a stepwise function $dist_{GED}$ is used, which incorporates a matching threshold θ with an Euclidean distance function $d(a_i, b_j)$, as per (4.3), to differentiate between matches and non-matches. If $d(a_i, b_j) < \theta$, then zero is assigned to $dist_{GED}$, otherwise $dist_{GED} = d(a_i, b_j)$, which is referred to as the mismatch penalty.

With the proposed GED algorithm a linear gap model is employed with gap penalty γ associated with each indel. Initial testing on spatial sequences demonstrated that the linear gap model on average produced better discrimination across the datasets than with other more complex gap models. This finding is consistent with gap model investigations in bioinformatics sequence alignment approaches (Gotoh, 1982).

The GED can be calculated using a dynamic programming matrix C of size $(|\mathbf{a}| + 1) \times (|\mathbf{b}| + 1)$ and iterating over i and j as shown in (4.4)-(4.7). The global distance score can be found at position $(|\mathbf{a}| + 1, |\mathbf{b}| + 1)$ within C.

$$C(0,0) = 0 \tag{4.4}$$

$$C(i,0) = \gamma i, \quad 1 \leq i \leq |\mathbf{a}| \tag{4.5}$$

$$C(0,j) = \gamma j, \quad 1 \leq j \leq |\mathbf{b}| \tag{4.6}$$

$$
\begin{aligned}
C(i,j) &= \min\{C(i-1,j-1) + dist_{GED}(a_i,b_j), \\
&\qquad C(i-1,j) + \gamma, \\
&\qquad C(i,j-1) + \gamma\} \qquad (4.7) \\
\text{where, } dist_{GED}(a_i,b_j) &= \begin{cases} 0 & d(a_i,b_j) < \theta \\ d(a_i,b_j) & d(a_i,b_j) \geq \theta \end{cases}
\end{aligned}
$$

From (4.4)-(4.7), if θ is large in relation to γ, the final alignment will favour matches over gaps. By favouring mismatches over gaps, GED typically results in shorter overall alignments as gaps add to the alignment length. If θ is small in relation to γ, the final alignment favour gaps over matches, resulting in larger alignment lengths.The pseudocode for the GED algorithm is provided in Algorithm 3 for further clarity.

The GED algorithm can be adapted for symbolic spaces by replacing $d(a_i, b_j)$ and θ in the matching conditions of the stepwise function $dist_{GED}(a_i, b_j)$ with $a_i = b_j$ (for matches) and $a_i \neq b_j$ (for non-matches). A positive constant ω is also required to replace the Euclidean distance as a mismatch penalty. Consideration must be made to the value of ω in relation to γ, as per θ. If $\omega < \gamma$ then one can expect more gaps to be introduced into the alignment in relation to mismatches, whilst if $\omega > \gamma$ then the opposite occurs and GED will produce more mismatches than gaps. The nature of the recognition task dictates the values of ω and γ. For example, if the task is to recognise spatial patterns with several significant gaps, then one would specify a small γ in relation to ω. Like LCSD, calculation of optimal GED parameters (in relation to recognition performance) are dataset dependent.

4.4 Evaluation of LCSD and GED

To evaluate the LCSD and GED approaches in relation to noise and recognition performance, spatial activity sequences from 10 activities (Dataset A) and 3 spatially similar activities (Dataset B) are utilised from Section 2.6. Each activity is captured at 10fps with the resulting sequence consisting of x, y coordinates, representing the activity path.Fig. 4.1 illustrates the innate temporal variation between the activities in dataset A. Similar variability was also evident in dataset B (not shown). Further investigation of the dataset sequences confirm that the innate variation is due to tracking system noise

Algorithm 3: Global Edit Distance (GED)

Data: θ, γ, **a**, **b**
/* *Initialise DP matrix C* */ ;
for $g \leftarrow 0$ **to** $|\boldsymbol{a}|$ **do**
  for $h \leftarrow 0$ **to** $|\boldsymbol{b}|$ **do**
    $C(g,h) \longleftarrow 0$;
/* *Initialise first row and column of DP matrix C according to linear gap penalty* */ ;
for $i \leftarrow 1$ **to** $|\boldsymbol{a}|$ **do**
  $C(i,0) \longleftarrow \gamma \times i$;
for $j \leftarrow 1$ **to** $|\boldsymbol{b}|$ **do**
  $C(0,j) \longleftarrow \gamma \times j$;
/* *Perform GED calculation on remainder of rows and columns in DP matrix*/ ;
for $i \leftarrow 1$ **to** $|\boldsymbol{a}|$ **do**
  for $j \leftarrow 1$ **to** $|\boldsymbol{b}|$ **do**
    if $dist_{GED}(a_i, b_j) < \theta$ **then**
      /* *Match between sequence elements* */ ;
      $aligned = C(i-1, j-1)$;
    else
      /* *Mismatch between sequence elements* */ ;
      $aligned = C(i-1, j-1) + dist_{GED}(a_i, b_j)$;
    /* *Insertion / Deletion (indel) in sequence - using linear gap penalty* */ ;
    $indel_a = C(i-1, j) + \gamma$;
    /* *Insertion / Deletion (indel) in sequence - using linear gap penalty* */ ;
    $indel_b = C(i, j-1) + \gamma$;
    /* *Find the minimum between match, mismatch and indels and set for* $C(i,j)$ */ ;
    $C(i,j) = min(aligned, indel_a, indel_b)$;

and intra-activity temporal variation.

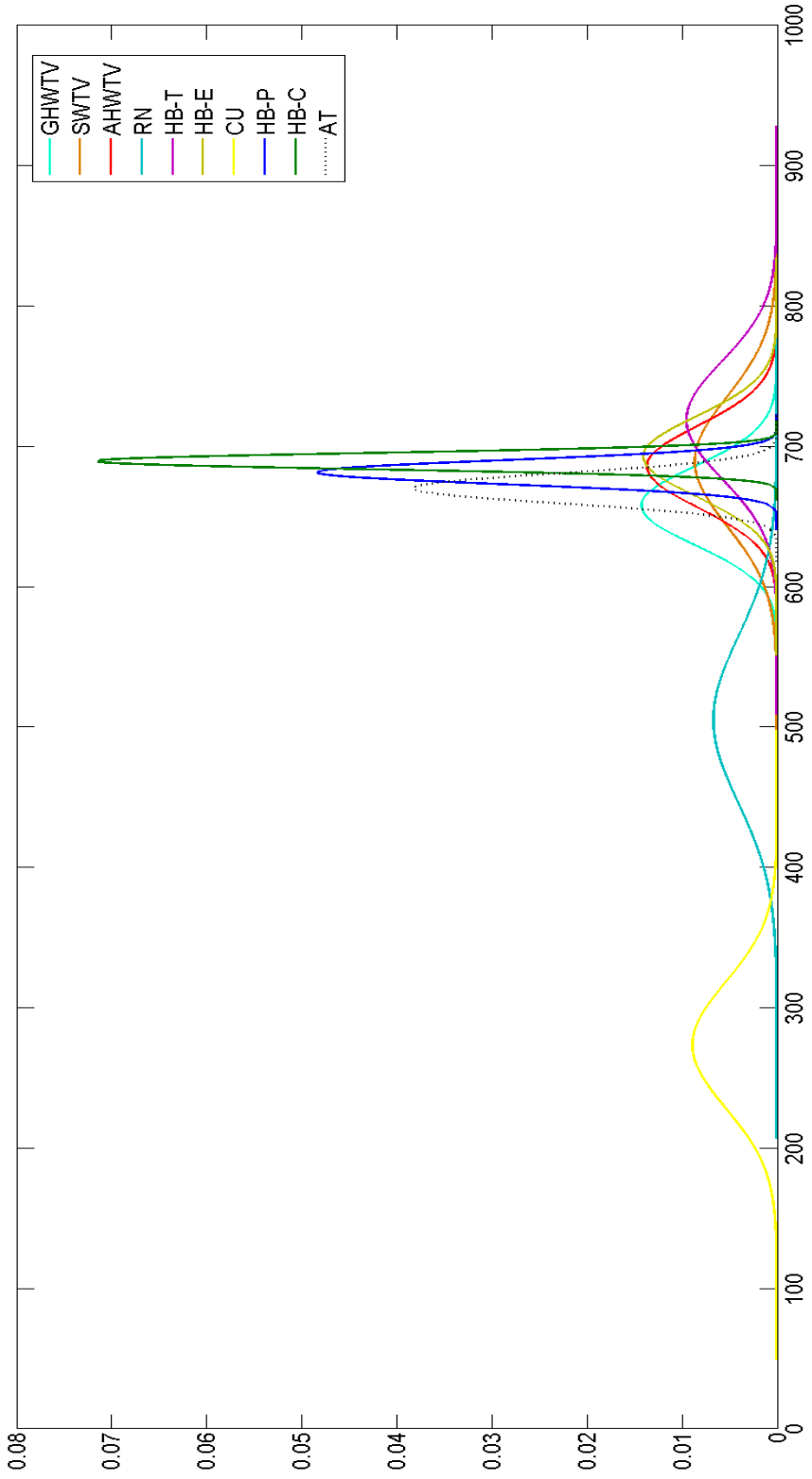


Figure 4.1: Activities exhibiting temporal variation (x axis represents the sequence length).

To provide a benchmark for the proposed LCSD and GED algorithms, the results are compared to the LCSS method of Vlachos *et al.* (2002a). In this approach the authors specify a distance threshold δ and a temporal matching threshold ϵ for matching spatial sequences. Here $\delta = 1.0m$, which is similar to θ, and ϵ is set to a large value to negate its' effect on the classification accuracy. This approach enables one to focus the empirical comparison on distance discrimination. To validate activity class recognition in the presence of temporal variation, the results are compared to DTW. A discrete HMM with fixed number of symbols $M = 156$, and an empirically determined number of hidden states N is also evaluated to ascertain performance against a generalised model. HMM's are trained and evaluated using discretised versions of the activity sequences as outlined in Section 2.6.5. For each data set, models with $N = 5$ to $N = 15$ hidden states were evaluated with $N = 5$ having the optimal recognition performance.

4.4.1 Parameter Selection

An empirical approach is adopted here to determine the optimal γ for GED. θ is not evaluated in this context as its values should be specified according to the recognition task. For instance, if the recognition task requires observed patterns to be strictly matched to those used in the training set, one would set θ to a small value. On the other hand if the matching constraint is relaxed and θ is set to a larger value, then one can expect to match more patterns and possibly obtain a higher degree of misclassification. The value of $\theta = 1.0$ was found to be a good compromise of precision and recall (5.2) for randomly generated training sets, obtained from datasets in Section 2.6. The fact that a similar value of θ was able to provide high precision and recall across the multiple datasets, results from activities having similar or overlapping spatial activity paths, similar durations, and/or consistent noise and errors from the tracking system. The value of $\theta = 1.0$ also coincides to the size of the discretised states used by HMM evaluations and is used throughout this thesis for the sequence alignment based approaches.

As specified in Section 4.3 γ is a linear gap penalty associated with insertion or deletion of one more trajectories or symbols in either sequence. Using a fixed θ, the issue of selecting an appropriate value of γ for the proposed GED algorithm is addressed. Selecting an optimal value for γ with a linear gap function is difficult in real world settings as its value can determine the degree of sequence deviation allowed (possibly correlating to

activity interweaving). Ideally γ, which is the penalty associated with insertions and deletions, should be the same as the penalty applied to sequence mismatches for a balanced outcome. In these experiments values of γ from 0.5 to 10.0 were evaluated in a cross-validation study to empirically derive an effective γ. For brevity we only show the resulting classification accuracies for the dataset B in Fig. 4.2; however the dataset A produced similar trends. For the given datasets, $\gamma = 1.0$ provided the maximum classification accuracies and thus was used in further experimentation.

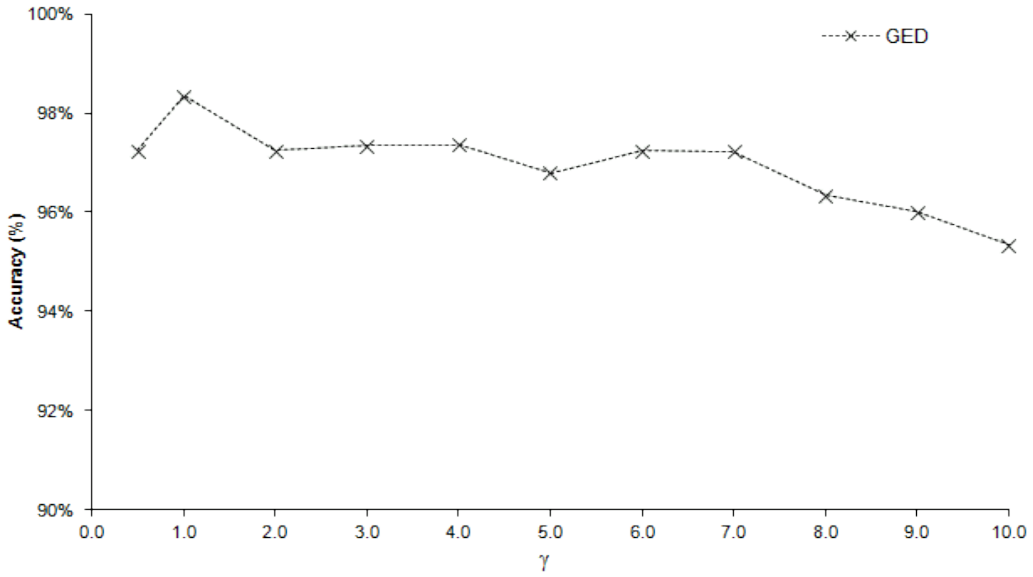
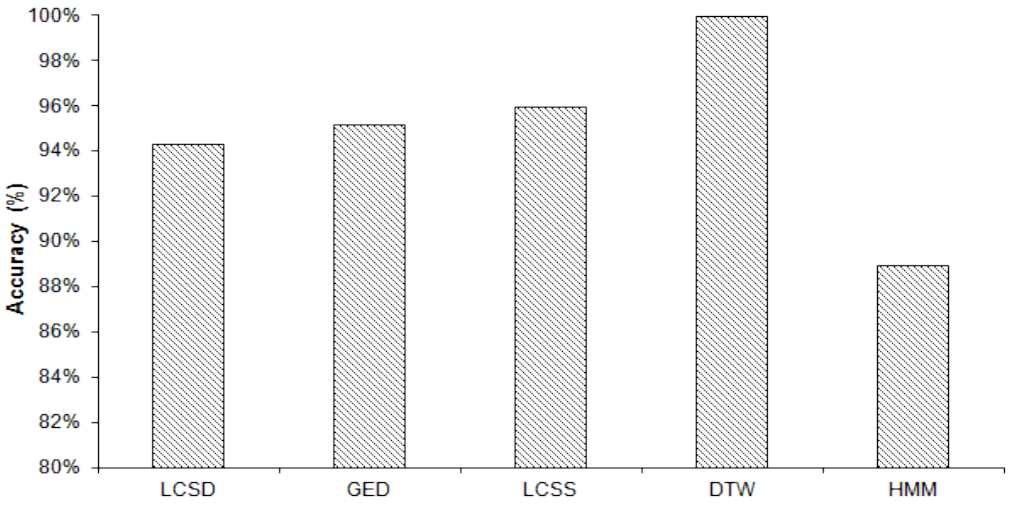


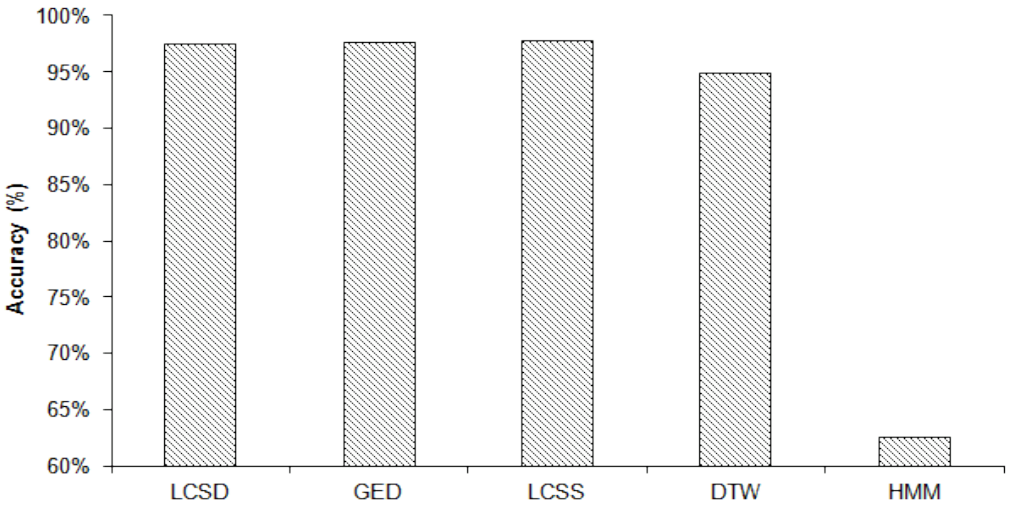
Figure 4.2: GED γ Parameter Optimisation.

4.4.2 Recognition Performance of LCSD and GED

In this section the recognition performance of the proposed LCSD and GED alignment approaches are evaluated using the parameters from Section 4.4.1, with datasets A and B, and cross-validation. Results are shown in Figures 4.3(a) and 4.3(b). Evaluation of the LCSD, GED algorithms demonstrated that the proposed algorithms are capable of accurate discrimination, outperforming the discrete HMM in both data sets. The HMM results from data set B were considerably lower than either the LCSD and GED further demonstrating the robust matching properties of these techniques. The poor



(a) Dataset A



(b) Dataset B

Figure 4.3: Recognition performance for LCSD, GED compared to LCSS, DTW and the HMM.

performance of the HMM in this instance may be attributed to the significant noise apparent in dataset B and/or the high degree of spatial overlap in the sequences preventing generation of a good discriminative model.

The DP-based approaches of LCSS, DTW, LCSD and GED did perform better than the discrete HMM, with DTW achieving higher accuracy with dataset A than either of the LCSD or GED approaches. Given low intraclass variations in spatial sequences, DTW is more accurate than the proposed approaches, as seen in Fig. 4.3(a). However, if spatial sequences do contain significant amounts of noise, DTW does not perform as well as the other DP-based algorithms such as LCSD and GED. This can be clearly seen by the lower accuracy achieved with DTW using dataset B in Fig 4.3(b). The observed decrease in DTW accuracy with spatial sequences containing significant amounts of noise is likely due to the algorithm minimising the distance across the whole alignment, thus taking into account the noise in the resulting distance score. The LCSS approach also exhibited similar spatial discriminatory performance to the proposed LCSD and GED algorithms as shown in Figures 4.3(a) and 4.3(b). LCSS is capable of robust recognition due to the underlying LCSS algorithm not taking into account any sequence dissimilarity. The fact that all sequence dissimilarity is ignored in LCSS is seen as a weakness of the approach, especially when quantifying spatially and temporarily similar, yet different activities.

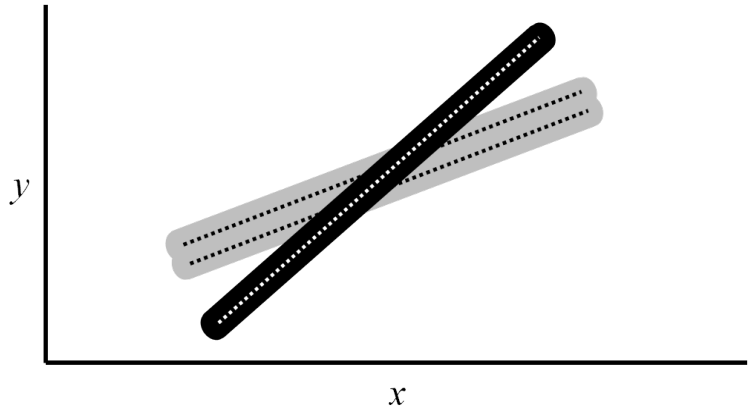
It is interesting to note that all the DP-based approaches evaluated in this experiment significantly outperform the discrete HMM. To exclude the possibility that the use of discretised sequences with the HMM is responsible for the lower observed accuracy, the same discretised activity sequences were applied to the one-dimensional versions of the LCSS, LCDS, GED and DTW algorithms. On average a less than 4% variance in accuracy was observed across the different approaches (results not shown), indicating that discretisation was not the cause. To ensure a sufficient number of states N were used for HMM learning and to ensure the models were not too generalised, the HMMs derived from $N = 5$ to $N = 15$ states were reevaluated and the accuracy reassessed. To ensure the results were significant, cross-validation was verified during the learning and inferencing phases. On average a 5% variation in accuracy was observed across the models, demonstrating that over-generalisation was not likely the cause of the HMM's observed lower accuracy with data set A. Over-generalisation is still a possibility with data set B, which exhibited a large degree of overlap between the three activity classes. One would expect that a larger number of states should capture the variation; however, even at $N = 15$ the accuracy was commensurate of the $N = 5$. An alternate hypothesis is that the HMM was performing as expected and that the DP-based approaches were just more sensitive at quantifying the distance (LCSD, GED, DTW) or similarity (LCSS) between activity classes. This increased sensitivity is implied when we look at the high resulting accuracy of the LCSD, GED, DTW and LCSS approaches across both data

sets in contrast to the HMM.

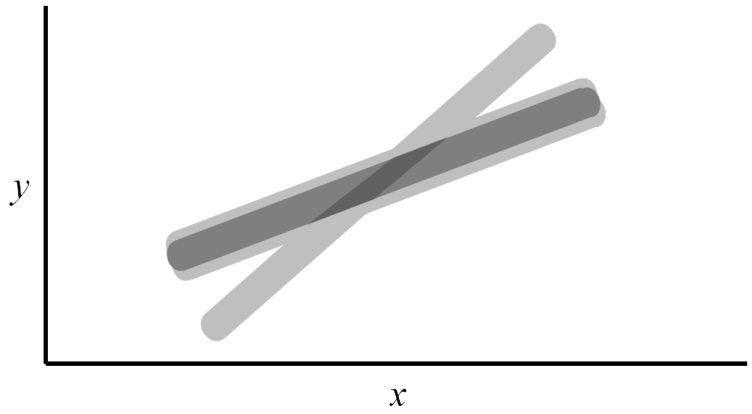
LCSD, GED, DTW and LCSS approaches are DP techniques that attempt to minimise the overall distance (or maximise the overall similarity) in an iterative fashion, based on optimal alignments of subsequences. This alignment occurs between a test sequence and a template sequence in relation to a scoring function that quantifies matches, mismatches, insertions, and/or deletions. Thus the final alignment and distance (or similarity) measure are highly dependent on the scoring function and its parameters. In these experiments we randomly select a set of template sequences, in this case 10, and use these to represent the activity and compare them to the remaining test sequences using a nearest neighbour approach. The scoring parameters such as θ and γ are empirically derived in accordance with user defined matching constraints to provide an effective level of segregation as explained in Section 4.4.1.

With DP-based approaches, each template sequence could be thought of as representing a separate warping alignment path within a multi-dimensional space, where the full set represents all the possibilities with which a test sequence could align (in whole or part). In contrast to the DP-based approaches, the HMM generalises its' template sequences into a singular alignment path where the symbol density is dependent on the degree of alignment space overlap (the more the variability, the less the overlap and the lower the density and vice versa). As DP-based approaches provide non-overlapping warping alignment paths for each template, they favour improved alignment and scoring with test sequences that are outliers in regards to the *average* sequence of a template set (assuming the template set has some variability). With the HMM, outlier sequences try to align to low density areas of the generalised alignment path, resulting in poor alignment probabilities and thus poor classification accuracy. This concept is demonstrated through Fig. 4.4, where the DP-based approaches form non-overlapping warping paths for 3 template sequences 4.4(a) [2 similar, 1 outlier], while the HMM forms a single generalised alignment model 4.4(b). If we apply a test sequence that is an outlier (similar to the sequence in black in Fig 4.4(a)) to both the DP-based and HMM approaches, we can see that the DP-based model would align well to the black template in Fig 4.4(a) achieving a strong alignment and corresponding similarity (or distance) score. In contrast the generalised HMM approach would align to the lower density alignment path (light grey) in Fig 4.4(b) resulting in a weaker alignment and similarity (or distance) score (as the outlier sequence has been observed less than the two similar sequences). This lower density alignment path in the HMM would correlate to smaller state transi-

tion probabilities across A and output probabilities B and is thus shown in a lighter grey colour. If we have activities with a degree of variability in their template sequences for a given activity set, we would thus expect sequence alignment approaches to outperform the discrete HMM. .



(a) Non-overlapping paths



(b) Single generalised alignment path

Figure 4.4: Conceptual representation of how DP-based techniques form non-overlapping warping alignment paths in contrast to the HMM which creates a single generalised alignment path.

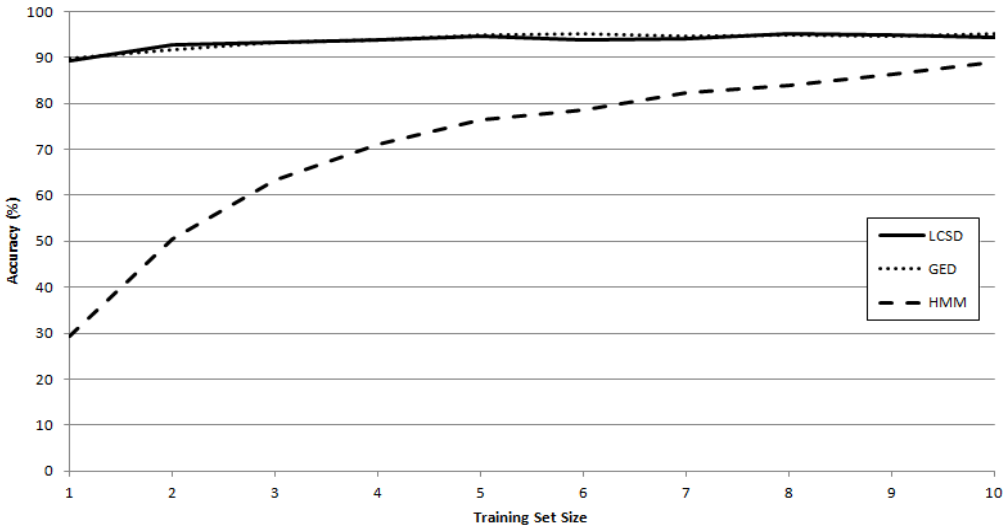
4.4.3 Training Set Size versus Recognition Performance

The training set size and its relationship to recognition performance is important as ideally one would like to obtain a high classification accuracy with the minimal number of templates representing an activity type. In these experiments the recognition performance is calculated for LCSD and GED algorithms, using the parameters obtained from Section 4.4.1 with the results compared to a discrete HMM. The recognition performance of the approaches with differing numbers of template sequences can be found in Figures 4.5(a) and 4.5(b).

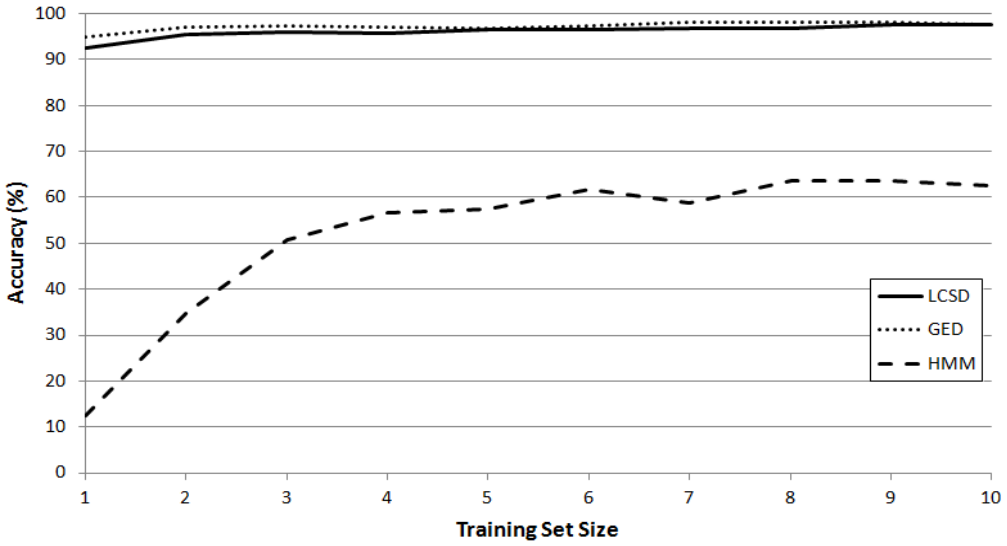
Through analysis of the recognition performance of the different approaches in Figures 4.5(a) and 4.5(b), LCSD and GED algorithms consistently achieve higher accuracy across the different training set sizes in comparison to the HMM. They also obtain a classification accuracy in excess of 89%, even with only one template in the training set. This contrasts the discrete HMM, which obtained a classification accuracy of only 88.9% with 10 training sequences. Overall, the HMM was observed to have significantly lower classification performance over both data sets using small training set sizes. This finding is consistent with the observation by Rabiner (1989) that discrete HMMs require large numbers of training sequences to produce good discriminatory models. The observed high accuracy of the proposed LCSD and GED algorithms with small numbers of templates demonstrates that sequence alignment based approaches do have high discriminatory properties when applied to a spatial recognition context. Furthermore, this discrimination characteristic is highly useful for recognising activities with limited exemplars.

4.4.4 Effect of Noise on Recognition Performance

In these following experiments the robustness of the LCSD and GED approaches are evaluated with dataset B and results contrasted to a discrete HMM (with $M = 156$, $N = 5$and models built with 10 sequences). Dataset B was selected due to the spatial similarity between the different activities (see 2.6), making the dataset more susceptible to misclassification with noise introduced from video tracking systems. To carry out the experimentation Gaussian noise is artificially incorporated with magnitudes up to 2.5 metres into each of the test sequences. Results of the evaluation are shown in Fig. 4.6.



(a) Dataset A



(b) Dataset B

Figure 4.5: Training set size versus accuracy (%) for datasets A and B.

Analysis of the results demonstrate that the proposed LCSD and GED algorithms are more resilient to noise than the HMM, as indicated by the lesser decrease in accuracy with the increased magnitudes of noise: $< 2\%$ decrease for the DP based approaches versus 12% for the HMM. Importantly, the 2D LCSD, GED and SW algorithms were able

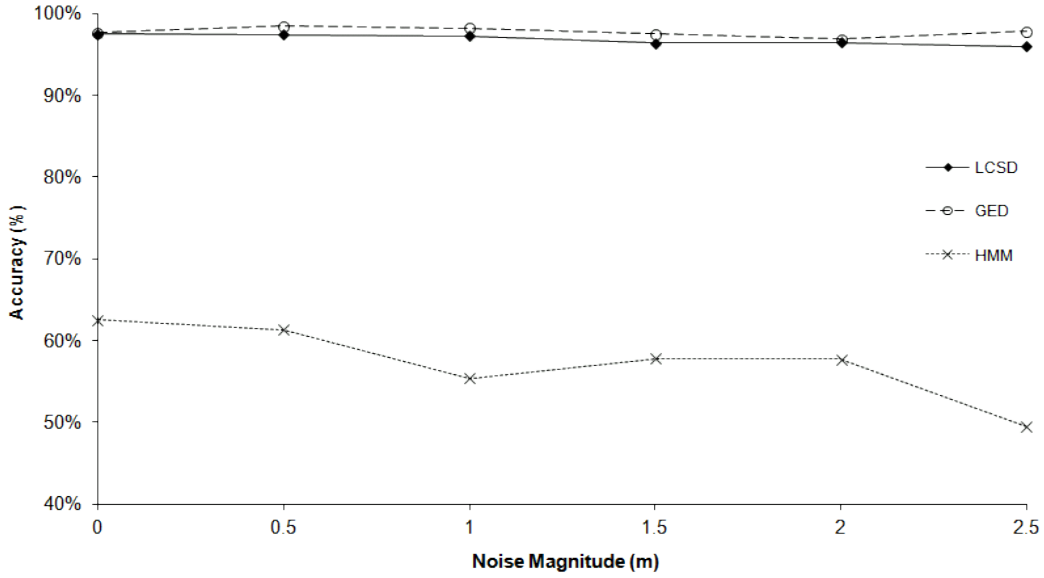


Figure 4.6: Noise magnitude versus accuracy for dataset B.

to maintain a classification accuracy of $> 96.0\%$ with the largest evaluated magnitude of noise, while the discrete HMM achieved only 49%. The maintenance of high accuracy across the different magnitudes of artificially introduced noise demonstrates that the LCSD, GED and SW algorithms are robust to noise and more so than the discrete HMM. This allows the algorithms to be applied successfully to similar pattern recognition tasks that have intrinsic noise issues.

4.5 Summary

In this chapter the LCSD and GED algorithms are proposed for accurate and robust use in recognising spatial activity sequences obtained from a smart home video tracking system. These algorithms are inspired by similar bioinformatics sequence alignment approaches that provide approximate matching with biological sequences exhibiting indels (expansion or compression) and substitutions due to natural and evolutionary pressures. The recognition performance of the LCSD and GED algorithms are assessed with spatial activity data and the results contrasted with existing approaches including LCSS,

DTW and the discrete HMM. Results show that the LCSD and GED techniques are more discriminatory than both DTW and the HMM for quantifying spatial sequences containing large amounts of spatial variability and noise.

Chapter 5

Recognising activity sequences in the presence of noise and temporal variation

In Chapter 4, LCSD and GED approaches were developed and shown to be robust to noise associated with spatial activity sequences in a smart home environment. This chapter further explores the challenge of recognising spatial activity sequences containing noise from video-based tracking systems in addition to recognising human activities with temporal variation. Techniques from DTW and the field of bioinformatics are used as inspiration for the investigation. The approach by Guo and Siegelmann (2004) attempts to address the noise and temporal variation issue in a music recognition context through the introduction of the Time-Warped Longest Common Subsequence (T-WLCS) approach; however, this symbolic approach is still sensitive to minor fluctuations in sequences due to the symbolic representation and simple scoring model that is employed.

Context is an important factor in the temporal aspect of human activities. The duration over which an activity is conducted can be affected by external contextual factors such as:

- The time of day, week, month or season. For example, activities conducted during the week are on average conducted faster, than the same activities conducted during weekends.
- Complexity of the activity. Complex activities can be achieved using different sequential orders of actions to achieve an end state, resulting in different average times. This can be illustrated when individuals try to assemble a piece of furniture without reading the instructions.
- Frequency with which the activity is conducted. Activities conducted frequently

are more homogeneous in how they are conducted and the durations over which they occur.

- Location. For example, making breakfast at work is typically shorter in duration than making breakfast at home.

The study of daily activities by Szalai (1972) further analyses and identifies the factors affecting human activities at a multi-national and population perspective. In an activity recognition application within a smart home environment, automatic activity recognition systems must be able to generalise human activity types with temporal variation. In this chapter, a bioinformatics-inspired template matching approaches, which are analogous to "time warping" in pattern recognition, are developed and evaluated for spatial activity recognition. Bioinformatics sequence alignment techniques are typically used to analyse DNA, RNA or protein sequences obtained through laboratory research. In this context exact matching is not always pertinent and it is more important to find approximate matches to a given pattern. The ability to recognise approximate matches is important as biological sequences are subject to change with natural or evolutionary pressures Crochemore and Rytter (2002).

The following chapter is organised as follows. Section 5.1 discusses an improvement on DTW which takes into account sequence alignment characteristics, referred to as Threshold Dynamic Time Warping (TDTW). This approach is capable of temporal activity recognition and is robust to noise introduced by camera tracking systems. A band DP approach is then outlined in Section 5.1.1 which improves the runtime of the TDTW approach. The experimental methodology and results to validate the effectiveness of the TDTW technique are provided in 5.2 respectively. Lastly, a summary is presented in section 5.3.

5.1 Threshold Dynamic Time Warping (TDTW)

This section describes a DTW-based approach that specifically deals with with temporal variation, yet is robust to intrinsic sequence noise. The original DTW algorithm is capable of accurate spatial sequence discrimination; however, the technique is sensitive to noise as it requires all elements of the sequences to be mapped to a corresponding

element(s) of the other sequence.

To address the robustness problem of DTW with spatial sequences and to allow to matching of activities exhibiting temporal variation, a new approach called Threshold DTW (TDTW) is explored. Similar to the LCSD and GED bioinformatics-inspired approaches, TDTW incorporates a threshold θ that specifies the maximum allowable Euclidean deviation for trajectory elements to match. One can conceptualise the θ parameter as a user specified window or buffer around a two dimensional template, as seen in Fig. 5.1 Choosing an appropriate value of θ for spatial sequence quantification

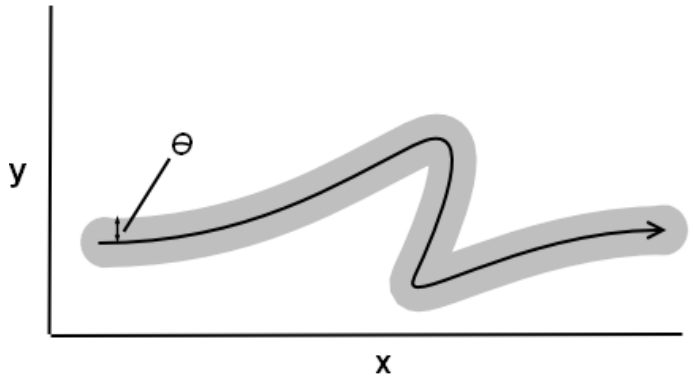


Figure 5.1: Windowing effect of θ on a two dimensional x, y coordinate space.

is application specific. However, if θ is too large, TDTW over generalises and matches dissimilar sequences and if θ is too small, then matching becomes more specific. In the special case when $\theta = 0$, TDTW reverts to DTW.

To reduce warping between the sequences in TDTW, thresholding of the local distance between trajectories is applied to the diagonal matching condition $C(i-1, j-1)$ of the DTW relation in (2.2). This forces more diagonal matches in the TDTW DP matrix, thus preventing minor warping with small changes in spatial position. The resulting formulation for TDTW using (4.1) is specified in (5.1) for $i = 2, \ldots, |\mathbf{a}|$ and $j = 2, \ldots, |\mathbf{b}|$, with C initialised according to (2.1) for $i = j = 1$, $1 < i \leq |\mathbf{a}|, j = 1$ and $i = 1, 1 < j \leq |\mathbf{b}|$.

$$C(i,j) = \begin{cases} C(i-1,j-1) & d(a_i,b_j) < \theta \\ \min \left\{ \begin{array}{l} C(i-1,j-1) \\ C(i-1,j) \\ C(i,j-1) \end{array} \right\} + d(a_i,b_j) & d(a_i,b_j) \geq \theta \end{cases} \tag{5.1}$$

The procedure to recover an optimal alignment for TDTW is similar to that of DTW. The following spatial sequences **a** and **b** are used to illustrate this procedure:**a** = $[(1.0, 1.0)(1.5, 1.5)(2.0, 2.0)]$ and **b** = $[(1.0, 1.0)(1.5, 1.5)(2.2, 2.2)(3.0, 3.0)]$. The completed DP matrix is generated using $\theta = 1.0m$ and is shown in Table. 5.1. Table. 5.2 illustrates a completed DP matrix generated using the traditional DTW approach, which is provided for comparison. One of the possibly many optimal warping paths

Table 5.1: TDTW C matrix using example sequences **a** and **b** with $\theta = 1.0$m.

	(1.0,1.0)	**(1.5,1.5)**	**(2.0,2.0)**
(1.0,1.0)	0.000	0.000	1.414
(1.5,1.5)	0.000	0.000	0.000
(2.2,2.2)	1.697	0.990	0.000
(3.0,3.0)	4.525	3.111	1.414

Table 5.2: DTW C matrix using example sequences **a** and **b**.

	(1.0,1.0)	**(1.5,1.5)**	**(2.0,2.0)**
(1.0,1.0)	0.000	0.707	2.121
(1.5,1.5)	0.707	0.000	0.707
(2.2,2.2)	2.404	0.990	0.283
(3.0,3.0)	5.233	3.111	1.697

for TDTW and DTW are represented in Tables. 5.1 and Tables 5.2 with underlined characters. Optimal warping paths can be derived by applying a traceback procedure to the DP matrices. An exemplar path is given in Fig. 5.2.

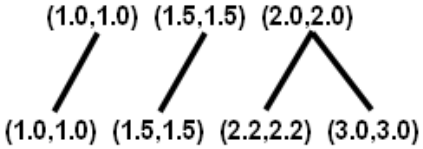


Figure 5.2: An optimal warping path for **a** and **b**.

5.1.1 Band DP constraint

To address the high computational complexity of DTW, global constraints have been applied to prevent excessive warping and to reduce superfluous DP calculations. Two of the more popular global constraints include the parallelogram and banding methods. The parallelogram DP constraint is applied to allow additional warping in the centre region of the sequences and away from the sequence termini Itakura (1975). As warping at the sequence termini is tightly constrained, the technique assumes that sequence beginning and end points can be accurately determined. The Sakoe-Chiba band was introduced in Sakoe and Chiba (1978) to provide additional warping at sequence termini for speech recognition. Band DP, a similar windowing approach was later proposed in Das (1982). Both approaches utilise a window length parameter to constrain DP calculations to a window with its centre running from the beginning of the DP matrix, corresponding to the beginning of the sequences, to the endpoint of the matrix for band DP or near the endpoint, such that the window encompasses the endpoint, for the Sakoe-Chiba band Sakoe and Chiba (1978).

The classical DTW approach (2.2) provides no restriction on the slope of the warping as it has no local continuity constraints, unlike the approach of Sakoe and Chiba (1978) and others. Selection of an optimal local constraint is application and domain specific and in the case of spatial activity recognition, investigations have confirmed that no restriction on the slope of the warping is optimal in relation to recognition performance. Unfortunately, the computational complexity of DP-based approaches make them difficult to apply in real life situations. Thus to prevent pathological warping of sequences and to reduce the computational complexity of DTW and TDTW, a band DP constraint is applied as specified in Das (1982). A band DP constraint restricts the possible warping paths in C by limiting the DP calculations to those within a window of size m running from $(1, 1)$ to $(|\mathbf{a}|, |\mathbf{b}|)$ of the DP matrix C (Fig. 5.3). As a result, only a portion of the DP matrix is calculated, reducing the runtime complexity of DTW and TDTW from $O(|\mathbf{a}||\mathbf{b}|)$ to $O(m|\mathbf{a}|)$.

Historically, the window size is set to 10% of the comparative sequence; however, Ratanamahatana and Keogh (2004b) has shown that different window sizes can produce higher accuracy depending on the data set.

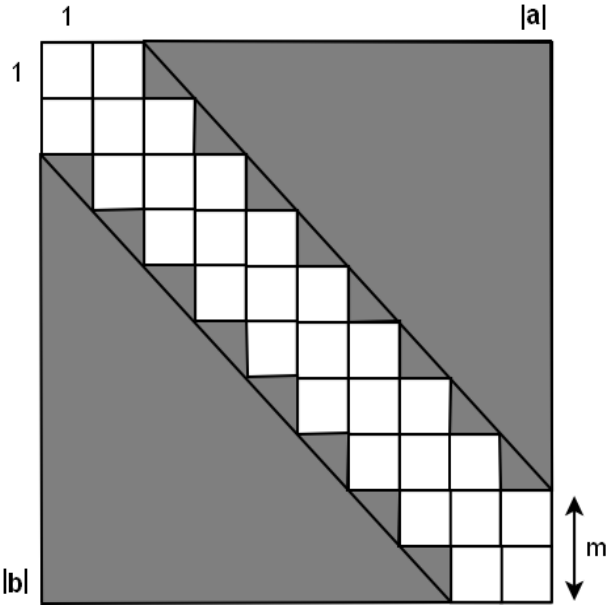


Figure 5.3: Application of a Band DP constraint with window size m.

5.2 Evaluation of Threshold Dynamic Time Warping (TDTW)

TDTW experimentation uses dataset B spatial sequences as the activities are spatially similar and therefore noise from tracking systems has a greater effect on classification accuracy. In the first experiment, the high discrimination ability and increased robustness claim of the proposed TDTW algorithm is validated by demonstrating how intraclass distances vary with artificially introduced noise for DTW and TDTW. More robust approaches exhibit a smaller variation in intraclass distance with noise compared to more susceptible algorithms. The intraclass distances of the same activities are then compared to the intraclass distances of different activities to provide evidence of the strong discriminatory capability of the proposed approach. Highly discriminate approaches have a marked separation between intraclass distances of correct and incorrect classes. The findings are confirmed with the given data set using threshold nearest neighbour (NN) classification and precision and recall statistics. Precision, defined in (5.2), measures the ability of a technique to correctly classify, whilst recall (5.3) measures the completeness of a technique's classification, that is the proportion of the true class test cases that were identified.

$$\textbf{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{5.2}$$

$$\textbf{Recall} = \frac{\text{True Positives}}{\text{Expected True Positives}} \tag{5.3}$$

To benchmark the resulting recognition performance of the proposed TDTW algorithm the results are also compared to the discrete HMM (Rabiner, 1989; Yamato *et al.*, 1992), with scaling (Rabiner, 1989) applied to both the model estimation and inferencing due to the length of the data set sequences. HMM models were trained using $M = 156$ symbols, $N = 5$ hidden states and 10 training sequences per activity, with A and B initialised as stochastic matrices of size $N \times N$ and $M \times N$, respectively. The number of hidden states $N = 5$ is selected from models derived from 5 to 15 hidden states, which were subject to cross-validation analysis with testing sequences. Baum-Welch parameter estimation is applied until convergence of the ratio of the average of the log-likelihoods between the current and previous iterations is achieved (less than 0.001). Cross-validation is again applied over 30 iterations to HMM model training and evaluation, where a new HMM is trained with a new randomised A and B with different training sequences, and the model performance evaluated with the remaining test sequences. To demonstrate that the discrete HMM is indeed sensitive to the amount of training data and that 10 training sequences produces the best performance in the following experiments, HMMs are trained with 1 to 10 random training sequences and evaluated (Fig 5.4). It is clear from Fig 5.4 that 10 training sequences produces a high recognition performance and a small error margin, and thus is suitable to be used in in later experiments for comparison with the DP-based techniques.

The cross-validation methodology is applied across the DP-based approaches whereby randomly generated training sets are compared against randomly selected testing sets. Following the accuracy and robustness experiments, classification accuracy is evaluated in relation to different band DP window sizes, in order to determine if band DP constraints are appropriate for DP-based problems in the spatial activity domain and secondly to determine an optimal window size for the given data set. To coincide with the discretisation of the two dimensional space for the discrete HMM approach, TDTW approaches used $\theta = 1.0$m, which also corresponds to the empirically derived optimal global threshold (as shown in 4.4.1).

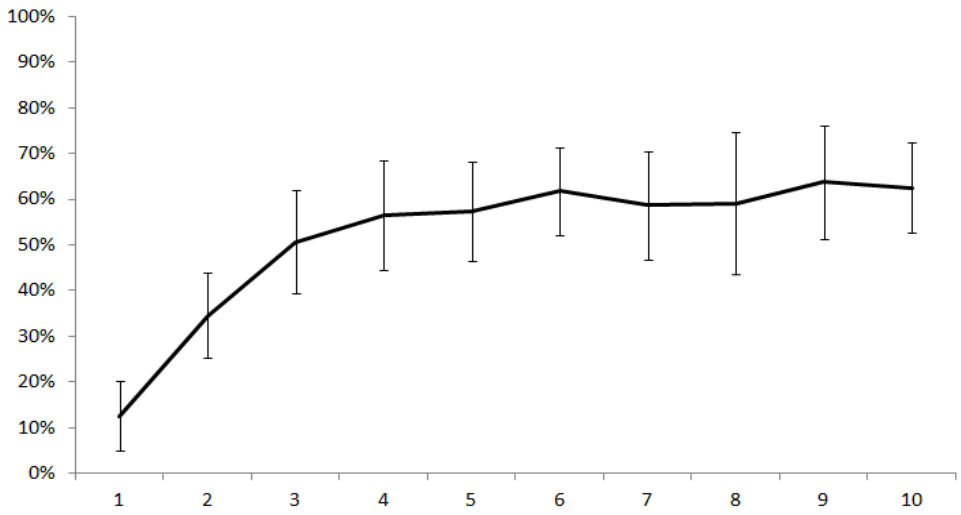


Figure 5.4: Recognition performance vs number of training sequences for the discrete HMM ($M = 156, N = 5$)

5.2.1 Effect of Noise on Recognition Performance

In order to evaluate the effect of noise on the recognition performance of the approaches, three separate experiments were conducted. In the first experiment, the intraclass distances of the three classes of activities are measured to determine how the average intraclass distance varies with noise. To carry out the experiment, we produced 30 randomly generated training sets of 10 sequences and compared each of these to the remaining 10 sequences of the same class. The results were then averaged across the 30 sets. The same methodology was applied for the sequences containing noise; however, Gaussian noise with a range of ± 3.0m was introduced to the testing sequences prior to distance quantification. The resulting difference in average distance between sequences containing noise and those without can be seen in Fig. 5.5.

The results from Fig. 5.5 show that TDTW is more resilient to noise than DTW, as the average intraclass distance increased by only half as much as DTW, in response to the artificially introduced noise. This is significant as one would typically apply a threshold for classification based on the intraclass distance. If an approach is highly sensitive to noise, the corresponding sequence distance will increase significantly with higher degrees of noise, thus reducing the recall statistics.

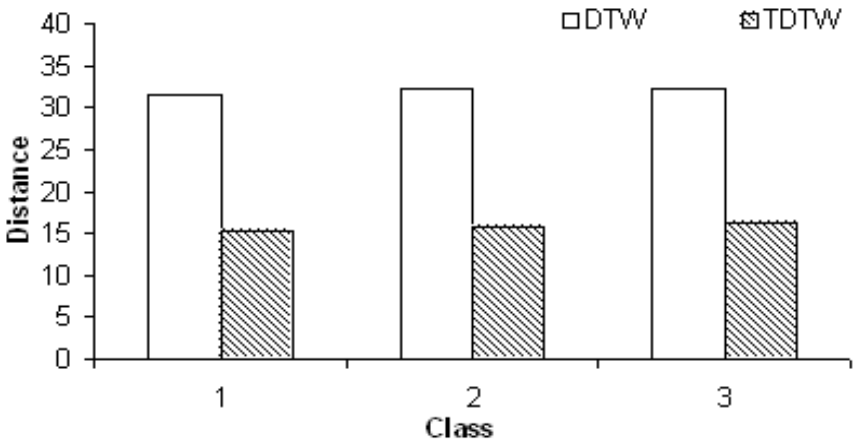
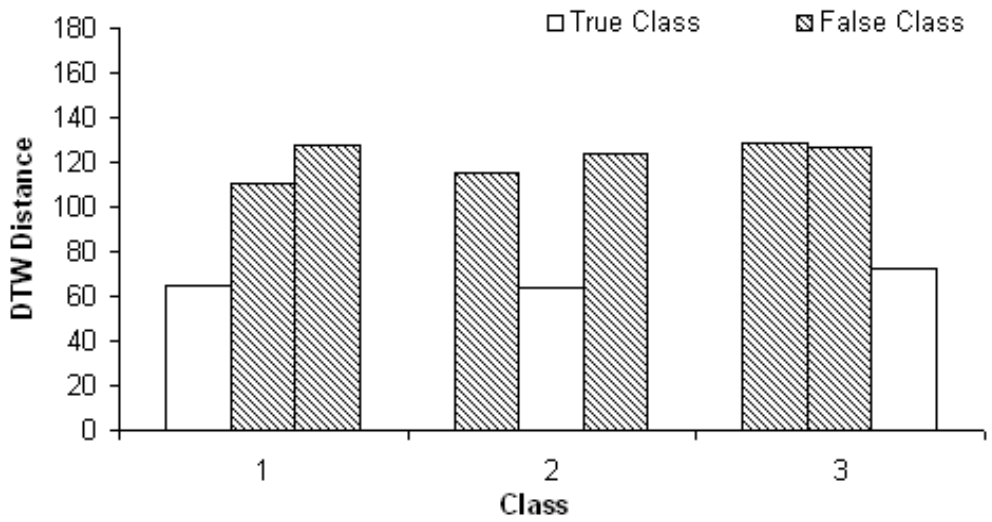


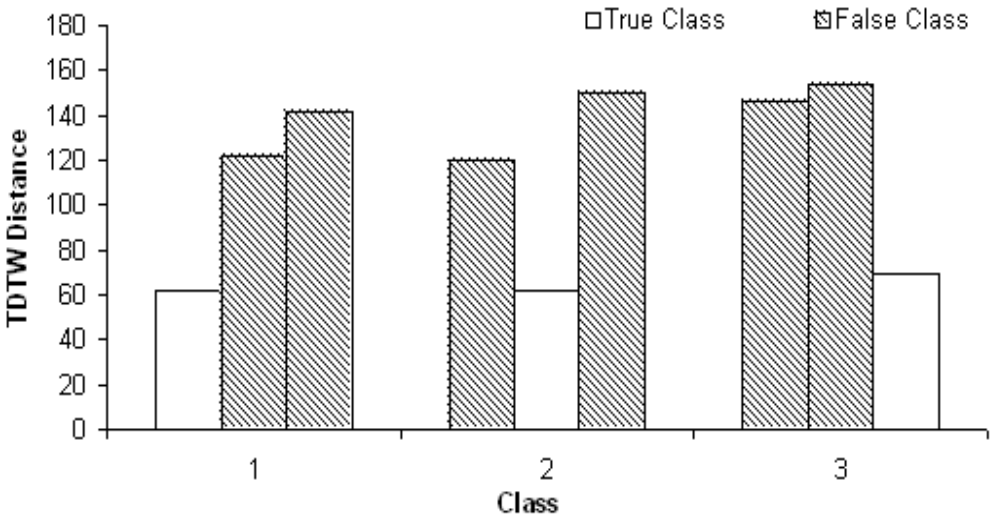
Figure 5.5: Average difference in DTW and TDTW intraclass distances (same activity) as a result of noise.

The second experiment examines the discriminatory ability of DTW and TDTW through measurement of the class distances between the three classes or types of activities. As mentioned previously, techniques that have a high discrimination ability exhibit good interclass separation between correct and all incorrect classes, thus decreasing the probability of misclassification. To evaluate the discrimination ability ten random training sequences were selected from each of the classes and the intraclass distances measured as per the first experiment. The training sequences of one class are then compared with the testing sequences (those sequences not used in training) of the other classes and the results averaged for each class over the 30 randomly generated sets. For instance, for class 1 in Figures 5.6(a)-5.6(d), the true class is the intraclass distance for class 1 sequences.

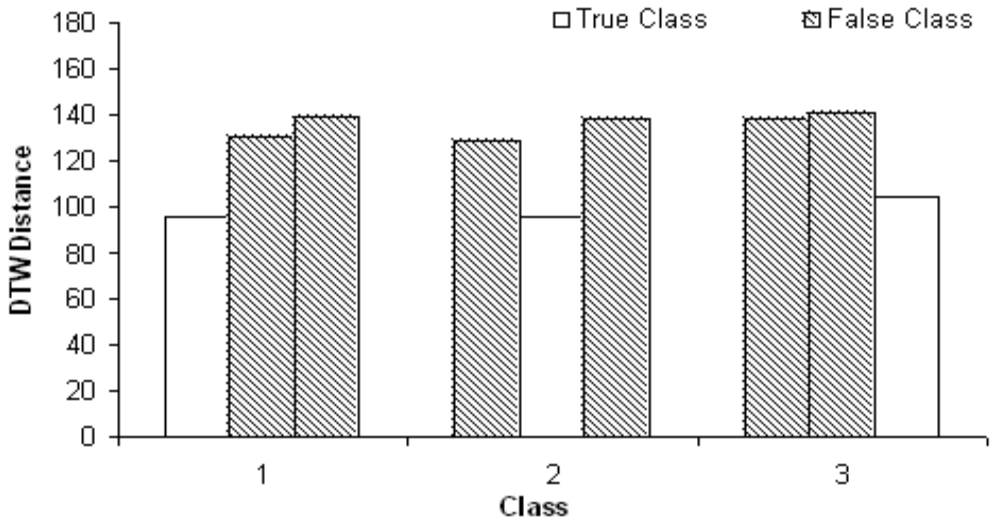
The two consecutive false class bars are the average intraclass distances for classes 2 and 3, which are calculated by comparing the testing sequences of class 2 and 3 with the training sequences from class 1. The true class for class 2 is the intraclass distance for class 2 sequences and the adjacent false class bars represent the average intraclass distances for classes 1 and 3 respectively, which are calculated by comparing the testing sequences of class 1 and 3 with the training sequences from class 2. The same experimental methodology was applied for the sequences containing noise; however, Gaussian noise with a range of ± 3.0m was introduced to the testing sequences prior to distance



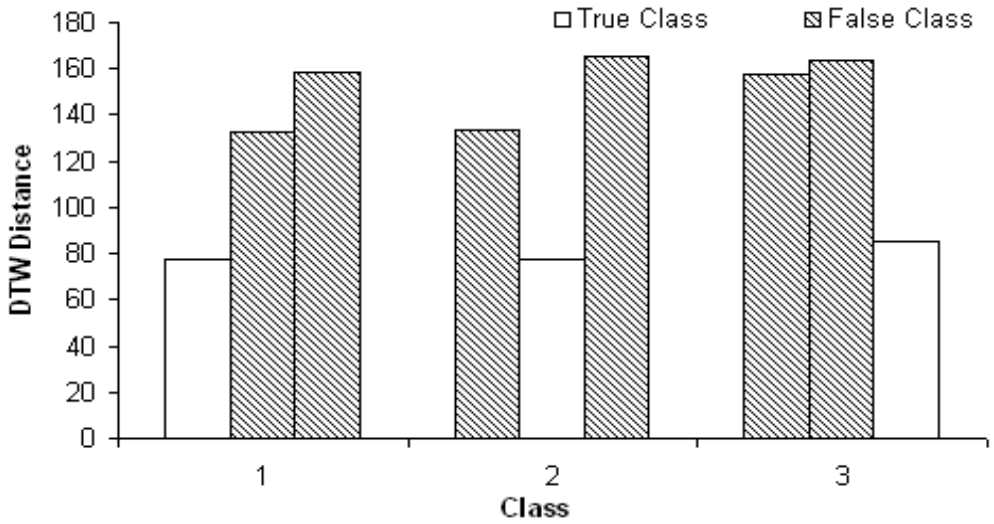
(a) DTW Class Distances



(b) TDTW Class Distances



(c) DTW Class Distances with a noise magnitude of ±3.0m



(d) TDTW Class Distances with a noise magnitude of ±3.0m

Figure 5.6: DTW and TDTW Class distances.

quantification.

The results in Figures 5.6(a)-5.6(d), show a significant difference in average class distance between true and false classes with the given data set. This demonstrates that both the DTW and TDTW approaches are able to provide high discrimination, even with sequences containing noise. Of significance is the fact that DTW had a smaller difference in class distance between true and false classes for both sequences containing noise and those without. The smaller difference in distance between correctly and incorrectly recognised classes indicates that DTW is more susceptible to misclassification than TDTW. If one compares the class distances of TDTW and DTW with respect to noise, both approaches respond similarly for the true classes. However, TDTW maintains the separation between true and false classes with the added noise, while DTW exhibits poorer separation, which will result in more misclassification and a lower recognition performance.

The third experiment validates the claim that TDTW is capable of accurate and robust spatial sequence recognition, using a threshold-based nearest neighbour (NN) evaluation methodology with ten training sequences or templates. To benchmark the DTW and TDTW results, discrete HMM models are generated for each activity class. Class thresholds are derived from the average intraclass distance plus two standard deviations for DTW and TDTW. HMM's are evaluated using two standard deviations due to the use of log likelihoods as a measure of sequence similarity. Recognition performance is ascertained using precision and recall statistics, according to (5.2) and (5.3). The results for the cross-validation study are shown in Table 5.3. The results show that the DP-based

Table 5.3: Precision and recall rates for threshold-based NN classification using DTW, TDTW and the discrete HMM.

Noise	DTW		TDTW		HMM	
(m)	Prec (%)	Recall (%)	Prec (%)	Recall (%)	Prec (%)	Recall (%)
0.0	96.3	94.7	98.3	97.3	59.7	56.9
3.0	96.1	88.3	98.3	91.3	56.0	56.1

approaches of DTW and TDTW produce significantly higher classification rates than the discrete HMM, even in the presence of noise. This justification has been explained previously in Section 4.4.2. It is also evident that TDTW exhibits increased classifica-

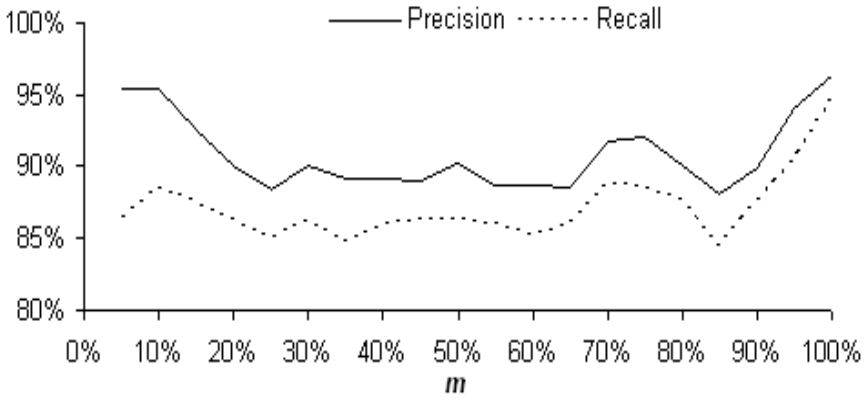
tion performance over DTW, especially in relation to precision. The higher classification performance of TDTW over DTW is attributed to the increased discrimination (larger interclass distances between true and false classes) resulting from inclusion of the matching region, provided by the threshold parameter θ. The inclusion of the matching region is shown to reduce the cumulative distance of similar sequences, whilst not affecting the distance to dissimilar ones; thereby, increasing the interclass distance between true and false activity classes and reducing the chance of misclassification.

In all three experiments, the DTW and TDTW approaches were capable of recognising activities that exhibited innate temporal variation resulting from video tracking artifacts as seen in Fig. 4.1. In TDTW, the incorporation of a real distance penalty and a θ-based stepwise function between sequence elements illustrated richer discrimination and robustness in relation to DTW. The high discrimination ability of both approaches show they are well suited for accurate recognition in a spatial activity recognition context. The results outlining the effect of increasing θ on precision and recall rates for TDTW are not included as it is well understood that increasing θ for the matching function will result in over-generalisation and increased misclassification as discussed in Section 6.1.

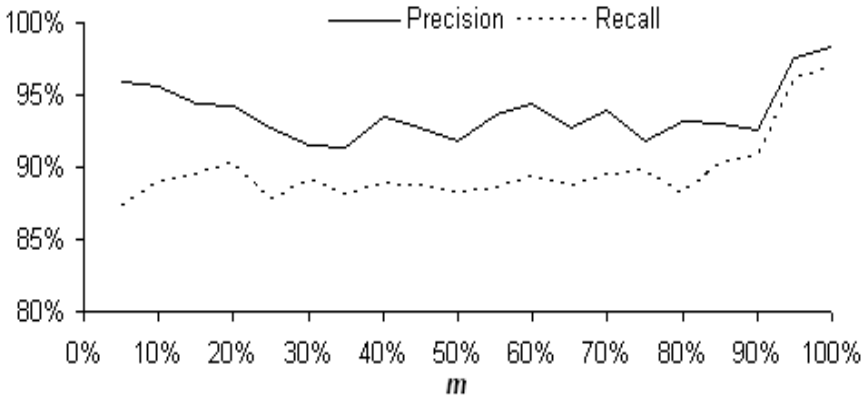
5.2.2 Effect of the Band-DP window size on Recognition Performance

As demonstrated in Ratanamahatana and Keogh (2004b) and Ratanamahatana and Keogh (2004a) introduction of a band DP constraint can improve classification accuracy and drastically reduce the computational complexity of DTW, particularly with small band sizes m. In the spatial recognition domain, applying such a constraint may improve the already high classification performance of TDTW, and in turn reduce the runtime complexity of the technique. In this experiment, values of m from 5% to 100% of the length of the comparative sequence are used for the cross-validation study. The runtime of each of the samples are also contrasted to the maximum runtime, corresponding to $m = 100\%$. Results are shown in Figures 5.7 and 5.8.

Our findings in Fig. 5.7 demonstrate that high classification accuracies can be obtained with band DP constraints and small band widths, where $m = 5\%$. Furthermore, as seen in Fig. 5.8, the use of small values of m dramatically reduce the runtime of DTW and TDTW by as much as 80-90%. The observed reduction in runtime with smaller values



(a) DTW



(b) TDTW

Figure 5.7: Band DP width m versus precision and recall.

of m is expected given the complexity analysis in section 5.1. Closer analysis of Fig. 5.7 also shows that the optimal window size, in relation to precision and recall, is observed at $m = 100\%$. The fact that precision and recall were maximised using the largest window size for both DTW and TDTW suggests that in the spatial activity domain band DP constraints decrease classification performance. In order to find out why large values of m produced the highest classification values we also analysed the captured data. Analysis revealed significant variation in the sequence length of sequences of the same class and likely arose due to the tracking system failing to consistently track the individual as they moved about their activity. As a result of the variation, sequences

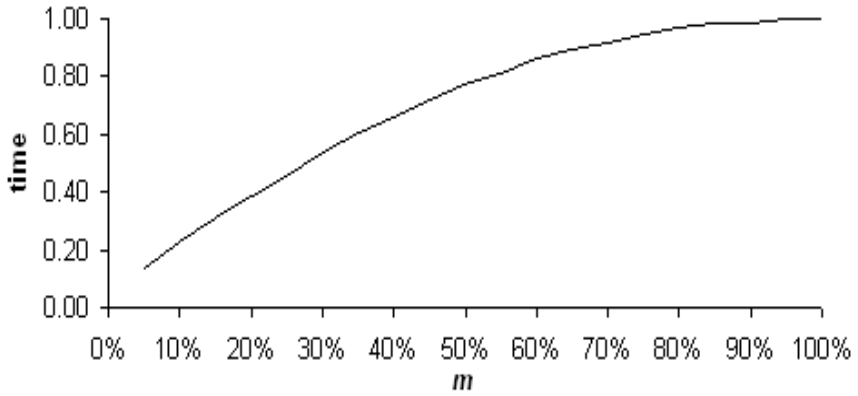


Figure 5.8: Band DP width m versus relative runtime.

required significant warping to temporally align, hence the large m observed. If the spatial sequences had been consistent throughout, then a lower optimal value of m should have been observed. In an ideal situation, with a tracking system capable of consistently tracking an individual, one can apply a band DP constraint in conjunction with a small window size to minimise runtime, whilst maximising recognition performance. With inconsistent sequences and band DP it is unlikely that small band widths will produce optimal classification; however, it is still possible to use the technique for efficient pruning of candidates prior to full quantification. For instance, one can apply the more efficient band DP technique with a small window size to select those sequences which show a high correspondence to a known pattern(s). Using this set of sequences, one can then apply the unconstrained DTW or TDTW technique for more accurate quantification, prior to classification.

5.3 Summary

In this chapter the TDTW algorithm is described and evaluated for accurate and robust use in recognising spatial activity sequences obtained from a smart home video tracking system. This approach is inspired by DTW and sequence alignment techniques for their approximate matching and time warping characteristics. TDTW is capable of

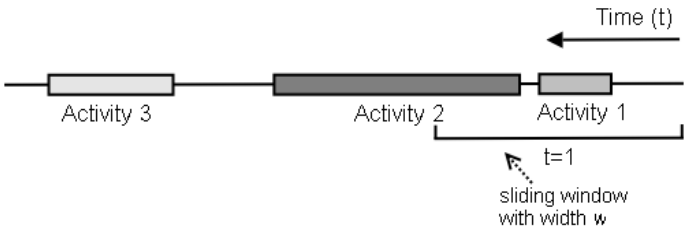
recognising activities with innate temporal variation and in the presence of noise generated by tracking systems. TDTW achieves its robustness through the introduction of a threshold parameter that denotes the maximum allowable Euclidean distance with which trajectories can match. By applying the thresholding concept to the diagonal matching condition of the DTW formulation, minor warping is suppressed, reducing sensitivity of the algorithm to noise. Evaluation of TDTW with a dataset of three spatially similar activities demonstrates that TDTW achieves higher classification than both DTW and the discrete HMM, even in the presence of Gaussian noise. Furthermore, runtime performance of the approach is shown to be dramatically improved, while retaining high classification performance, through the use of band DP constraints with small window sizes.

Chapter 6

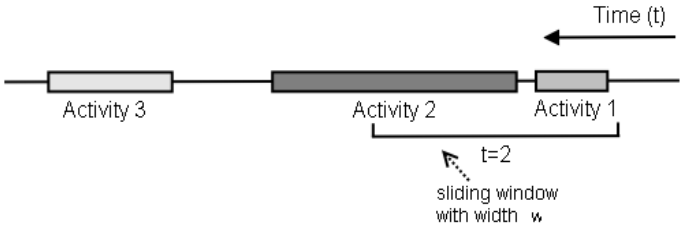
Recognising Embedded Activities within Spatial Sequences

In Chapters 4 and 5 the problem of recognising activities with intrinsic noise and temporal variation was addressed using techniques inspired by sequence alignment and DTW approaches. In this chapter the problem of recognising embedded activities is explored within continuous spatial sequences obtained from an online video tracking system. In the context of spatial activity recognition, online data streams can contain activities and non-activity subsequences, corresponding to movement between activities and deviations from known activity paths. In order to isolate individual spatial activity sequences for qualification or quantification with activity models or templates one can either use a sliding window, with width w on the buffered stream, or segment the data stream using "signature" based approaches. Fig. 6.1(a) - 6.1(c) describes the process of online recognition using a sliding window approach. Sliding window approaches compare the fixed length window sequence to pre-stored templates or models to quantify similarity prior to classification, as shown in Fig. 6.1(d). Segmentation techniques recognise and extract embedded activities through location of activity boundaries, prior to similarity quantification and classification. Segmentation of continuous data is not a new problem and has been addressed previously in domains such as speech and gait recognition. Unfortunately, in the spatial activity domain segmentation is more difficult as activity boundaries are not so obvious.

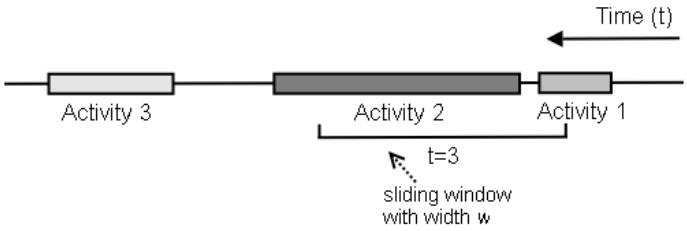
Few methods have been proposed to specifically deal with online activity segmentation. In the methods of (Bobick and Ivanov, 1998) and (Ivanov and Bobick, 2000a) a sliding window is applied to an observed sequence to allow for inferencing with low level HMMs. The observed sequence is then labelled accordingly. Like other sliding win-



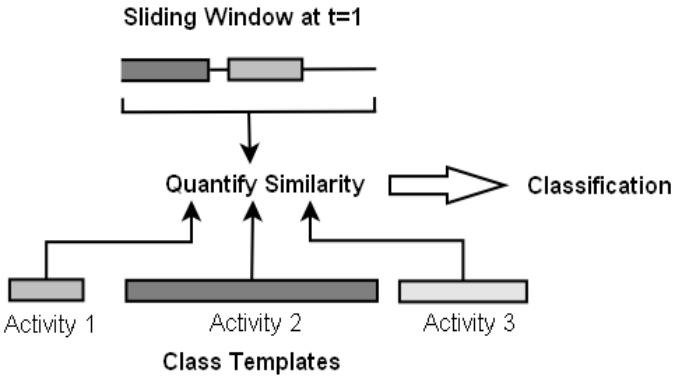
(a) Sliding Window at $t = 1$



(b) Sliding Window at $t = 2$



(c) Sliding Window at $t = 3$



(d) Classification Process

Figure 6.1: Online Spatial Activity Recognition using a Sliding Window.

dow approaches, the performance of this technique is sensitive to the specified window size. Another segmentation approach is given by (Peursum *et al.*, 2004), where observed sequences are segmented and classified using HMMs trained with manually labelled activity sequences. During classification, the probability of a sequence having a particular label is determined and through calculation of the probability at each time instance, the boundaries of the activities can also be found.

To apply any of the above activity segmentation methodologies in isolation, for segmentation of a continuous data stream, is problematic. This is because the segmentation components of the approaches are intertwined with the recognition capabilities. Therefore, one must still adopt a sliding window approach to identify embedded activities, particularly if one wishes to use unrelated sequence matching techniques for similarity quantification. Given this constraint, two issues relating to sliding windows need to be addressed. The first of these is the window size w. If one assumes that activities are conducted over a similar duration, and in ideal tracking conditions (such as in controlled indoor environments) then it is appropriate to use a window size corresponding to the length of the longest activity. Realistically, the window size must be set to some value larger than the longest activity length, taking into account a feasible increase in possible activity duration. With an appropriately sized sliding window (set to the length of the longest activity or average length ± 2.5 standard deviations), the second issue relates to locating an activity within that window sequence as shown in Fig. 6.1(d). Quantifying window sequences poses a problem for classification as the corresponding sequences can contain additional subsequence elements, which are not part of an embedded activity. These superfluous elements can in turn reduce the probability of an activity occurring in relation to a learnt model or increase the aligned distance between a class template. Even if temporal variation is consistent amongst activities, discrepancies in captured sequence length still occur due to tracking systems failing to consistently track objects. Some possible reasons for the failure result from occlusions, lighting variation, deficiencies in background subtraction techniques and geometric modelling limitations.

Techniques like the HMM (Rabiner, 1989) take the whole window sequence into account when calculating the probability of an observed sequence belonging to a given model. As a result, the superfluous elements decrease the resulting sequence probability, particularly if they have estimated symbol probabilities close to zero. Dynamic time warping (DTW) (Sakoe and Chiba, 1978) and similar global sequence alignment approaches such as edit distance with real penalty (ERP) (Chen and Ng, 2004) and edit distance on

real sequence (EDR) (Chen *et al.*, 2005) are also susceptible to superfluous sequence elements. This occurs as the techniques attempt to minimise the distance across the entirety of both the known and observed sequences, taking into account the additional distance from the superfluous elements. Sequence similarity algorithms based on LCSS (Vlachos *et al.*, 2002b,a) address this global limitation by ignoring superfluous elements in the observed sequences. Unfortunately, the techniques also allow significant deviations in a pattern, which can lead to incorrect classification as shown in 4.2.

In order to recognise known spatial activity patterns and to address the above mentioned deficiencies, the Smith-Waterman (SW) local alignment approach is used as inspiration to develop a two dimensional spatial activity recognition approach that is highly discriminative and can be applied to accurately and inaccurately segmented activity sequences. The derived SW algorithm is provided in Section 6.1, with nomenclature given in 4.1. A more efficient SW formulation for online recognition, called Online SW (OSW), is also outlined in Section 6.2. The developed SW approach is evaluated in regards to recognition performance with sliding windows containing activities, accurately segmented activity sequences and activity sequences containing Gaussian noise. The OSW approach is also evaluated with manually segmented activity sequences as well as embedded sequences from an online tracking system. To benchmark the classification performance of OSW, the approach is compared to DTW and a discrete HMM. Both SW and OSW evaluations can be found in Section 6.3. Lastly, a summary is provided in Section 6.4.

6.1 The Smith-Waterman (SW) Approach

Similarity-based sequence alignment techniques like Needleman-Wunsch global alignment typically have distance counterparts as the negative penalties assigned to non-matching sequence elements can be replaced by a positive penalty in the distance form. SW is distinct in that has no distance counterpart(Waterman, 1995). This is because the algorithm uses negative similarity in conjunction with the extra zero to terminate poorly matching subsequence alignments, which can't be mimicked in a distance based approach.

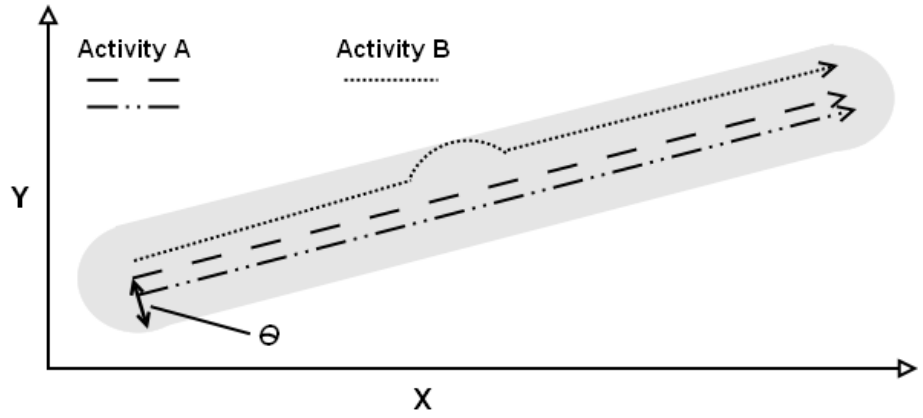
In order to apply local alignment to two dimensional spatial sequences several modifications are required of the original SW algorithm. An Euclidean matching function

$d(a_i, b_j)$ and corresponding matching threshold θ are firstly introduced as per LCSD, GED and TDTW. If the Euclidean distance between the symbols a_i and b_j is less than the matching threshold θ then a match results and a positive score is attributed, that is $s(a_i, b_j) = \alpha$. If the Euclidean distance is equal to or larger than the matching threshold θ, a real penalty is assigned indicating non-matching symbols. The real penalty $-d(a_i, b_j)$ is also based on the Euclidean distance (4.1). This approach penalises those symbols that are spatially further apart more than those symbols that are spatially closer together, which is appropriate for this context. The real penalty $-d(a_i, b_j)$ is then applied to the stepwise function $s(a_i, b_j)$. The use of a real penalty function rather than a constant has been shown in initial research (results not shown) to improve the discrimination capability of the SW algorithm.

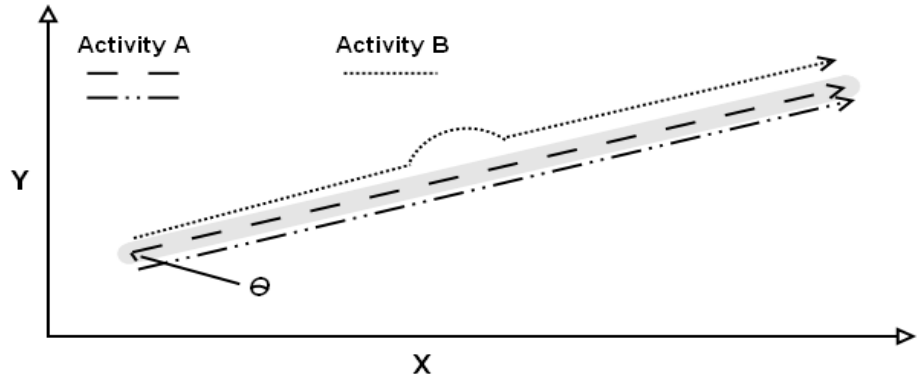
Like LCSD, GED and TDTW, choosing an appropriate matching threshold θ requires consideration of the level of specificity of the matching. If the specified θ of an x, y coordinate space is too large, the technique over generalises and matches dissimilar sequences as depicted in Fig. 6.2(a). This condition may be required when trying to identify new types of activities and determining the level of spatial similarity they have in regards to known activities. If θ is small, then matching becomes highly specific (Fig. 6.2(b)), preventing recognition of similar sequences and thus reducing the recall statistics. Decreasing values of θ can be advantageous when misclassification rates are high to improve discrimination.

As described in Section 2.4.3, gap scores are typically a function of the gap length l, denoted by $g(l)$. A linear gap model, where $g(l) = -l\gamma$, has equal weighting for gaps (caused by noise or activity variation). In this linear gap model it is assumed that the probability of a gap occurring in a sequence is the same anywhere along the spatial sequence. A linear gap model was found in preliminary investigations to be the most consistent in regards to recognition performance across the spatial activity datasets.

In this approach, alignments are dependent on the values of the gap penalty γ, match cost α and the matching threshold θ. If γ is larger in relation to the average mismatch penalty, which is dependent on θ, mismatches are favoured over gaps, producing shorter, more compact alignments. The opposite occurs when γ is smaller than the average mismatch penalty. In relation to the match cost α, one doesn't want $\alpha \gg \gamma$ or the mismatch penalty otherwise SW ignores mismatches and gaps and therefore behaves similar to LCSS.



(a) θ too large, Activity B is recognised as A



(b) θ too small, Alternate Activity A sequence not recognised

Figure 6.2: The affect of θ on spatial activity recognition.

To calculate the similarity of two spatial sequences **a** and **b** using the proposed SW based approach one simply applies (6.1)-(6.3) to the DP matrix C for $i = 0, 1, \ldots, |\mathbf{a}|$ and $j = 0, 1, \ldots, |\mathbf{b}|$ and finds the maximum value in C. At each $C(i, j)$ where $i, j \neq 0$, four choices (match or mismatch, gap in **a**, gap in **b** or start a new subsequence) are evaluated similar to SW with the choice corresponding to the maximum similarity value being selected for each $C(i, j)$. The match or mismatch score at each $C(i, j)$ is derived using $s(a_i, b_j)$ as previously described, while the gap scores for the sequences are derived using the linear gap model. If a negative similarity score results from $C(i-1, j-1)+s(a_i, b_j)$, $C(i-1, j)+\gamma$ and $C(i, j-1)+\gamma$, due to poor subsequence correspondence, then the fourth

option of starting a new subsequence, represented by zero, is selected as the maximum.

$$C(i,0) = 0, \quad 0 \leq i \leq |\mathbf{a}| \tag{6.1}$$
$$C(0,j) = 0, \quad 0 \leq j \leq |\mathbf{b}| \tag{6.2}$$

$$\begin{aligned} C(i,j) &= max\{C(i-1,j-1) + s(a_i,b_j), \\ &\qquad C(i-1,j) - \gamma, \\ &\qquad C(i,j-l) - \gamma,\ 0\} \end{aligned} \tag{6.3}$$
$$\text{where, } s(a_i,b_j) = \begin{cases} \alpha & d(a_i,b_j) < \theta \\ -d(a_i,b_j) & d(a_i,b_j) \geq \theta \end{cases}$$

To show how the SW algorithm works and the procedure to recover an optimal alignment, the following example spatial sequences are used as seen in Fig. 6.3. The completed

$\mathbf{a} = [(1.0,1.0),(2.0,2.0),(3.0,3.0)]$
$\mathbf{b} = [(0.0,0.0),(1.0,1.0),(2.0,2.0),(4.0,4.0),(5.0,5.0),(6.0,6.0),(1.0,1.0),(2.0,2.0),(3.0,3.0)]$

Figure 6.3: Example sequences **a** and **b**

C matrix generated with $\theta = 1.0m, \alpha = 1.0$ and $\gamma = 1.0$ is shown in Table. 6.1. One

Table 6.1: SW C matrix using example sequences **a** nd **b**.

		(1.0,1.0)	**(2.0,2.0)**	**(3.0,3.0)**
	0.00	0.00	0.00	0.00
(0.0,0.0)	0.00	0.00	0.00	0.00
(1.0,1.0)	0.00	1.00	0.00	0.00
(2.0,2.0)	0.00	0.00	2.00	1.00
(4.0,4.0)	0.00	0.00	1.00	0.59
(5.0,5.0)	0.00	0.00	0.00	0.00
(6.0,6.0)	0.00	0.00	0.00	0.00
(1.0,1.0)	0.00	<u>1.00</u>	0.00	0.00
(2.0,2.0)	0.00	0.00	<u>2.00</u>	1.00
(3.0,3.0)	0.00	0.00	1.00	<u>3.00</u>

of the possibly many optimal local alignment paths obtained from the DP matrix C is shown in Table 6.1 with underlined characters. The consequent optimal local alignment

is shown in Fig. 6.4. From Table 6.1 a second local alignment can also be formed between the subsequences [(1,1)(2,2)(3,3)] and [(1,1)(2,2)(4,4)]. However, the mismatch between trajectories (3,3) and (4,4) (given $\theta = 1m$) reduces the similarity to only 0.59, which is sub-optimal in comparison to the local alignment in Fig. 6.4.

```
                                                (1,1) (2,2) (3,3)
                                                  |     |     |
(0,0) (1,1) (2,2) (4,4) (5,5) (6,6) (1,1) (2,2) (3,3)
```

Figure 6.4: An optimal SW alignment using sequences **a** and **b**.

6.2 Online Smith-Waterman (OSW)

In an online recognition context where you have a stream of data and you apply a sliding window over that stream, a naive approach for determining SW similarity is to calculate the full DP matrix for each sliding window, resulting in an $O((w+1) \times (|C_i|+1))$ complexity for each window, where $|C_i|$ is the length of the class template sequence i. With these applications, window size w is typically set to the length of the longest training sequence. A more computationally efficient method is proposed here for online recognition applications, referred to as the online Smith-Waterman (OSW) approach. In this formulation, the online system uses n DP matrices of size $(w+1) \times (|C_i|+1)$, where n is the number of class templates. Each of the $i = 1, 2, \ldots, n$ DP matrices are initialised by calculating the SW score using (6.1)-(6.3) with the initial buffered window sequence of size w positioned at time t_0 to t_{w-1}. The initial window sequence comparison is then made against the corresponding class template C_i, according to Fig. 6.5. The n SW scores generated by the DP initialisation process are stored in memory for comparison to subsequent SW scores of sliding windows, where $t > 0$, and for comparison to predetermined classification thresholds θ_i for class templates C_i. The row corresponding to the maximum SW score in the DP matrix is also retained for each DP matrix to ensure the current maximum remains in the matrix as the window slides during incremental updating. The rationale for this is explained later.

Following initialisation of the n DP matrices and to process further window sequences, the concept of incrementally calculating DP matrix rows from prior subalignments is explored. This concept is based on the fact that optimal DP alignments are derived

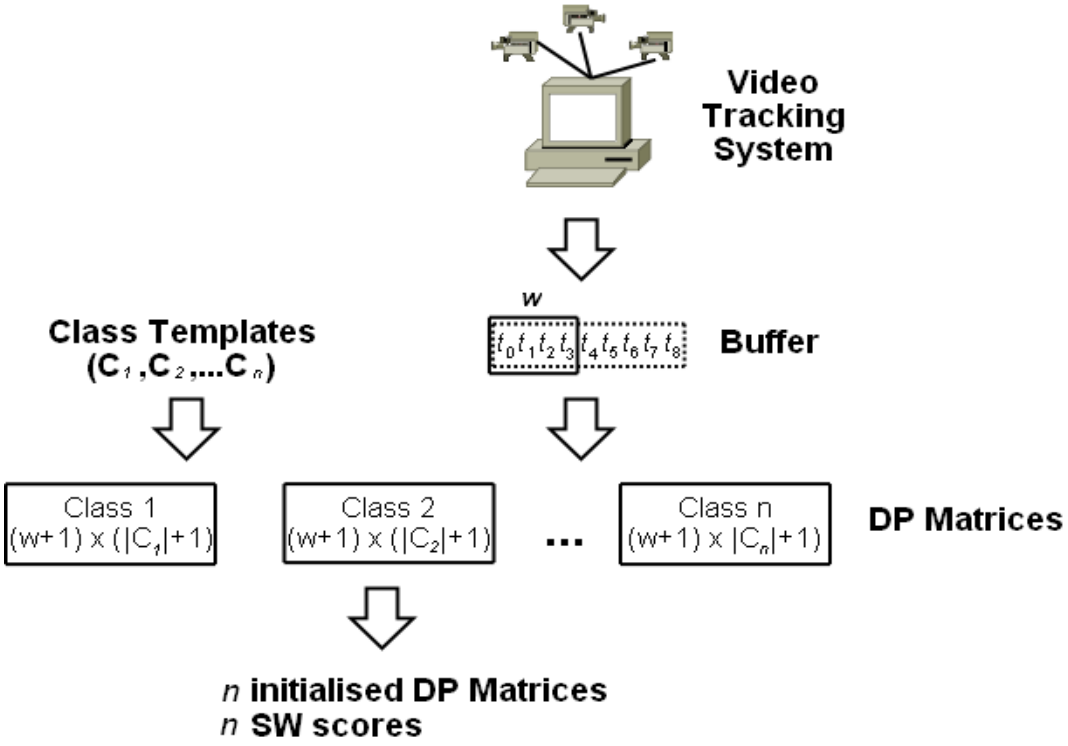


Figure 6.5: Initialisation of the DP matrices using OSW.

incrementally from subalignments that are also optimal. In order to illustrate the incremental derivation of optimal alignments, the following sequences $\mathbf{a} = [(1,1),(2,2)]$ and $\mathbf{b} = [(1,1),(2,2),(3,3)]$ are used, with $\theta = 1$,$\alpha = 1$ and $\gamma = 1$. The first row of the resulting DP matrix is initialised using (6.1), as shown in Table 6.2(a). The second row of the DP matrix (Table 6.2(b)), which corresponds to the optimal alignment between $\mathbf{a}$ and $\mathbf{b}_1 = [(1,1)]$, is then calculated using (6.3), with the first column of the row initialised to zero according to (6.2). The third row of the DP matrix, seen in Table 6.2(c)), is calculated by comparing $\mathbf{a}$ with $\mathbf{b}_2 = (2,2)$ according to (6.3). This extends the optimal alignment to between $\mathbf{a}$ and $\mathbf{b}_{1:2} = [(1,1),(2,2)]$, as the optimal alignment between $\mathbf{a}$ and $\mathbf{b}_{1:2} = [(1,1),(2,2)]$ comprises the optimal alignment between $\mathbf{a}$ and $\mathbf{b}_1 = [(1,1)]$, obtained through calculation of the second row of values. To calculate the final row in the DP matrix (Table 6.2(d)) a similar methodology is applied.

Table 6.2: Incremental Derivation of DP Matrices

(a) Step 1

	(1,1)	(2,2)
0	0	0

(b) Step 2

		(1,1)	(2,2)
	0	0	0
(1,1)	0	1	0

(c) Step 3

		(1,1)	(2,2)
	0	0	0
(1,1)	0	1	0
(2,2)	0	0	2

(d) Step 4

		(1,1)	(2,2)
	0	0	0
(1,1)	0	1	0
(2,2)	0	0	2
(3,3)	0	0	1

With incremental derivation one can add new rows to the end of each C_i DP matrix and along the window sequence axis (the vertical axis in the example), and calculate only the values of the row according to the new online element, the class template C_i and (6.3). This requires $O(|C_i|+1)$ time for each window, in comparison to the $O((w+1)\times(|C_i|+1))$ time of the naive approach. To prevent the DP matrices from growing in size as new rows are added and to allow a traceback procedure to be carried out on the DP matrix (to find segmentation beginning and end points with a sufficiently large window size w), the first rows of the DP matrices are removed prior to new rows being added. This also

means that the current max value *RowMax* is decremented for each DP matrix. In the event that *RowMax* designates rows that have been deleted from the DP matrix, a search of the matrix is conducted to find the new maximum score and *RowMax* is consequently updated. It is assumed that exemplar sequences will always be within (in most part) the window w, as the window size is derived from the maximum of the class sequence lengths, hence the requirement to find a new optimal subsequence when the *RowMax* is no longer in the DP matrix.

Initialisation and the functionality of OSW is outlined in algorithms 4 and 5 for further clarity. In these algorithms, OSW is applied to a window sequence w and a class template sequence c, using a DP matrix DP.

Algorithm 4: Online Smith-Waterman (OSW) Initialisation

Input : DP, SW parameters θ, α and γ
Output: calculated DP matrix, SW_{score}, $RowMax$
/* *DP is a matrix of size $(w+1) \times (c+1)$ and initialised as per (6.1) and (6.2)* ;
$t = 0$;
$SW_{score} = 0$;
$RowMax = 0$;
/* ***Step 1*** *- Calculate the full DP matrix at $t = 0$, determine the optimal subsequence score and row containing that score* */ ;
/* *Calculate SW values for w and c sequences as per (6.3)* */ ;
for $i \leftarrow 1$ **to** $|\mathbf{w}|$ **do**
    for $j \leftarrow 1$ **to** $|\mathbf{c}|$ **do**
        /* *Find the subsequence option that scores highest* */ ;
        $DP(i,j) = max\{DP(i-1,j-1) + s(w_i, c_j), DP(i-1,j) - \gamma, DP(i,j-1) - \gamma, 0\}$;
        /* *Check if alignment score for $DP(i,j)$ is higher than the previous recorded max as we go* */ ;
        if $DP(i,j) \geq SW_{score}$ **then**
            /* *Update score and row which contains highest score* */ ;
            $SW_{score} = DP(i,j)$;
            $RowMax = i$;

With OSW, optimal subsequences corresponding to an embedded activity within a sliding window (assuming w is large enough), can exist anywhere within the DP matrix, as the beginning and ends of the DP matrices are constantly changing. Therefore, to ad-

Algorithm 5: Online Smith-Waterman (OSW)

Input: calculated DP matrix, SW_{score}, $RowMax$

/ **Step 2** - Slide the window of width $|w|$ over the steam to $t+1$ */* ;

while *not at end of stream* **do**

/ slide window by one */* ;

$t \leftarrow t+1$;

$\mathbf{w} \leftarrow w((t-|\mathbf{w}|):t)$;

/ **Step 3** - Delete row 0 of DP to make room for next subsequence calculation at $t+1$ */* ;

$DP = DP(1:(|\mathbf{w}|+1))$;

/ **Step 4** - Decrement $RowMax$ and check if its still in DP. If not, search for new SW_{score} and $RowMax$ */* ;

$RowMax = RowMax - 1$;

/ If row containing maximum SW value is no longer in DP */* ;

if $RowMax \leq 0$ **then**

$SW_{score} = 0$;

/ Search for new SW_{score} and $RowMax$ in DP */* ;

for $i \leftarrow 1$ **to** $|\mathbf{w}|$ **do**

for $j \leftarrow 1$ **to** $|\mathbf{c}|$ **do**

if $DP(i,j) \geq SW_{score}$ **then**

/ Update score and row which contains highest score */* ;

$SW_{score} = DP(i,j)$;

$RowMax = i$;

/ **Step 5** - Calculate SW for new row at $t+1$ corresponding to $DP(|\mathbf{w}|+1)$;*

/ Go to the last row of w */* ;

$i = |\mathbf{w}| + 1$;

/ Calculate SW for the row */* ;

for $j \leftarrow 1$ **to** $|\mathbf{c}|$ **do**

/ Find the subsequence option that scores highest */* ;

$DP(i,j) = max\{DP(i-1,j-1) + s(w_i, c_j), DP(i-1,j) - \gamma, DP(i,j-1) - \gamma, 0\}$;

/ Check if alignment score for $DP(i,j)$ is higher than the previous recorded max as we go */* ;

if $DP(i,j) \geq SW_{score}$ **then**

/ Update score and row which contains highest score */* ;

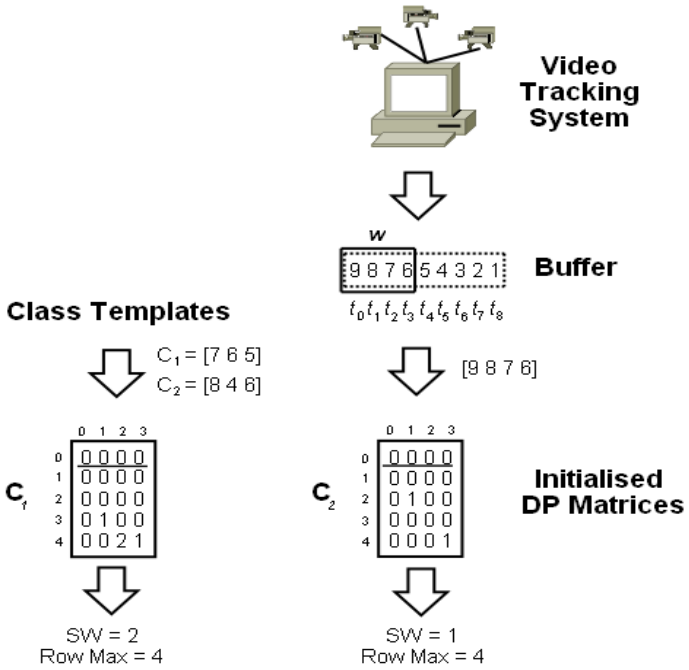
$SW_{score} = DP(i,j)$;

$RowMax = i$;

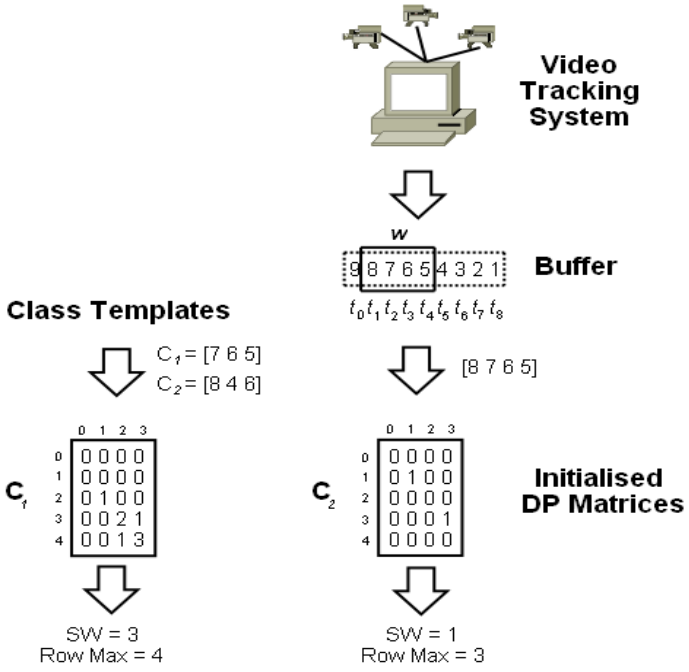
dress this issue the current maximum similarity value and its row (*RowMax*) are stored for each of the i matrices. The row position corresponds to the end of the optimal local alignment and is used as the origin for a traceback procedure, while the maximum similarity values are used in thresholding and classification. For clarity we show the online DP matrix update procedure with 1D sequences in Figures 6.6(a) and 6.6(b), with a gap penalty, mismatch penalty and match cost of one. In Fig. 6.6(a) the windowed sequence [9876] at t_0 to t_3 is compared to the class templates C_1 and C_2 using (6.1)-(6.3), thereby initialising the DP matrices. As the sliding window is moved to the next window at t_1 to t_4 (Fig. 6.6(b)), the first underlined rows of the DP matrices in Fig. 6.6(a), are deleted and a new row is added to the end of the matrices. For each of the new rows at $j = 4$, we apply (6.3) using the previous rows values at $j = 3$, the class template C_i and the new element in the new sliding window, that is five in the given example, in order to derive the new SW local alignment. The row-by-row update procedure is repeated for further sliding window sequences. If an observed spatial sequence does not fit within the specified window size w, possibly due to an activity taking longer than expected (e.g. watching TV), the DP matrices retain the similarity scores of the previous matches and thus a similarity score can still be determined. Unfortunately, full segmentation can no longer occur as the beginning point of the observed sequence would have been deleted to allow calculation of the new buffer elements.

In specific applications, OSW can be modified to decrease space requirements and improve segmentation performance. For example, if segmentation and recovery of an optimal alignment is not necessary the space requirements of the OSW DP matrices can be significantly reduced by utilising only the last row of the matrices. This is possible as new elements only require the last row for the current SW calculation. Additionally, the *RowMax* variables are no longer required.

Some spatial activities like watching TV have inconsistent sequence lengths, thus making the selection of a suitable window size w difficult. If w is sufficiently large to deal with the highly variable sequence lengths, the efficiency of the OSW algorithm deteriorates rapidly. To maintain computational efficiency the window size is set to the length of the longest template sequence and only a single row of previously calculated SW values are proposed with which to perform the current SW calculation. Additionally, a traceback threshold ϑ is required for triggering the traceback procedure. To carry out the traceback another data stream buffer Buf_{past} containing κ previous window elements is necessary to store past elements. As new buffer elements are retrieved, a new row of the DP



(a) DP Initialisation with Window Sequence t_0 to t_3



(b) Online Updating with Window Sequence t_1 to t_4

Figure 6.6: OSW recognition with a window size of w=4.

vectors are calculated using the previous row and designated template C_i. The maximum value in that row SW_{max} is then determined and the value compared to the traceback threshold ϑ. If $SW_{max} < \vartheta$, the next element in the buffer is added to Buf_{past}, the previous DP vector is deleted and the process continues again with a new buffer element. However, if $SW_{max} \geq \vartheta$, indicating that a activity has been identified, a traceback and SW recalculation procedure is initiated to locate the beginning and end points of the activity.

6.3 Experimental Results

For the evaluation of the SW and OSW approaches, a 12 activity dataset (dataset C (2.6) is utilised. The dataset comprises activities such as making toast, having breakfast, washing the dishes and watching television with 20 sequences captured per activity. Discretisation of the sequences are carried out for the discrete HMM evaluation whereby x, y trajectories are mapped to a sequence of unique integers u, where $u \in U$ and $U = 1, 2, 3, \ldots, 72$. In the following experiments dataset C is divided into training and testing sets. Training set sequences are used to empirically determine optimal algorithm parameters and for use as class templates in testing. To quantify the recognition performance of the algorithms with the testing sets, a cross-validation methodology with threshold-based nearest neighbour (NN) classification is adopted. In this approach each experiment utilises 30 randomly generated training sets for evaluation, from which the mean of the 30 test results is used for analysis. Thresholds are derived and incorporated for each activity to determine whether an activity occurs. A recognition threshold is necessary as it is unrealistic to learn all activities that may occur in a given environment and secondly to assume that one of these activities are always occurring.

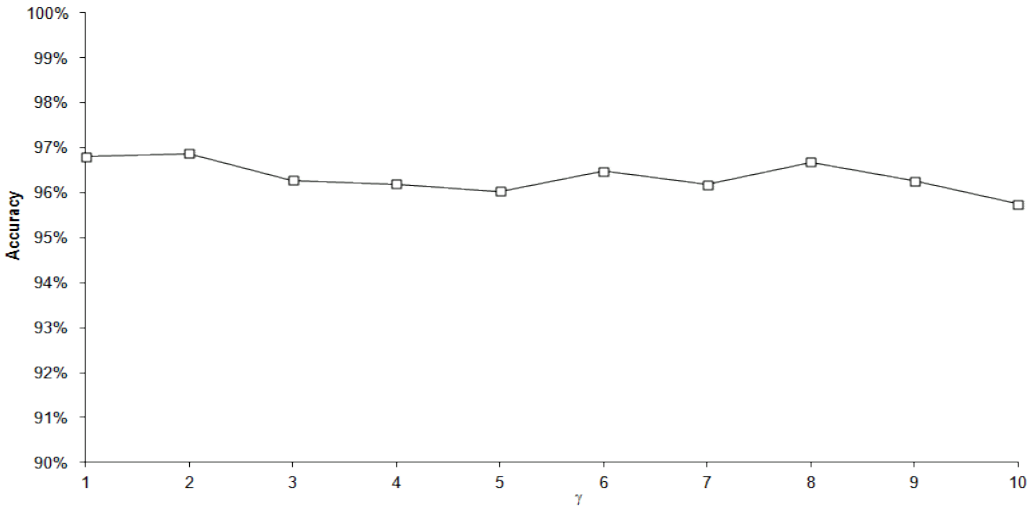
For the DTW benchmark comparison the symmetric algorithm defined in (Sakoe and Chiba, 1978) is used with no local or global constraints. Optimal SW and HMM algorithm parameters are empirically derived from the accurately segmented training data as shown in Section 6.3.1. Using the optimally derived parameters the proposed algorithms are evaluated in relation to accuracy and robustness, and results contrasted to the DTW and HMM spatial activity recognition approaches.

6.3.1 Parameter Selection

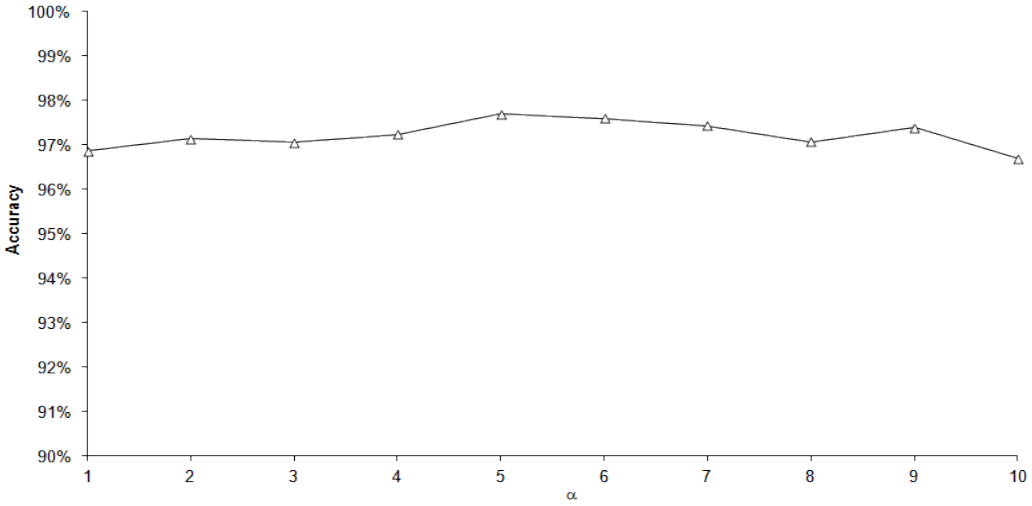
With accurately segmented activity sequences the optimal SW sequence alignment parameters are determined empirically to maximise accuracy, with the exception of θ which is set according to the required recognition task as per LCSD (4.2), GED (4.3) and TDTW (5.1). Throughout the experimentation, $\theta = 1.0$m to coincide with the size of the symbolic states used in HMM discretisation and the empirically derived optimal global threshold (as shown in 4.4.1).

Using a fixed θ the issue of selecting an appropriate value of γ is addressed for the proposed SW algorithm similar to GED in 4.3. As specified in Section 6.1 γ is the linear gap penalty associated with insertion or deletion of one more trajectories in either sequence. To minimise the effect of the match cost α during the γ evaluation α is set to a constant and values of γ, where $\gamma = 1, 2, \cdots, 10$, are used with cross-validation and NN classification in order to find an optimal value. The results with the given data set are shown in Fig. 6.7(a). From Fig. 6.7(a), a value of $\gamma = 2.0$ provided the maximum classification accuracy with larger values producing only marginally worse performance. Using $\gamma = 2.0$, an optimum value was determined for the match score α, where $\alpha = 1, 2, \cdots, 10$. Results are shown in Fig. 6.7(b) with $\alpha = 5.0$ producing maximum classification accuracy. Therefore, the following experiments used SW parameters of $\theta = 1.0m$, $\gamma = 2.0$ and $\alpha = 5.0$. It it interesting to note that recognition performance does not appear to be sensitive to the values of the different parameters (excluding θ), as evidence by the small changes in recognition performance with changing parameter values. The uniform recognition behaviour with different parameters has also been noted with other datasets in 2.6, indicating that it may be possible to choose a set of SW parameters without having to empirically determine the optimal parameter set, and expect a near optimal recognition performance.

To contrast the SW and OSW approaches, discrete HMMs with $M = 72$ and an empirically determined number of hidden states N were evaluated with dataset C. To ensure adequate training of the HMMs the number of iterations of the Baum-Welch estimation algorithm were limited by a threshold (< 0.001) applied to the ratio of the average of the log-likelihoods between the current and previous iterations. In order to find the optimum N in relation to accuracy, HMMs were generated with $N = 5, 6, \cdots, 15$ hidden states and evaluated with the training data in conjunction with NN classification. Results are



(a) γ evaluation.



(b) α evaluation.

Figure 6.7: Empirical SW Parameter Optimisation.

shown in Fig. 6.8).

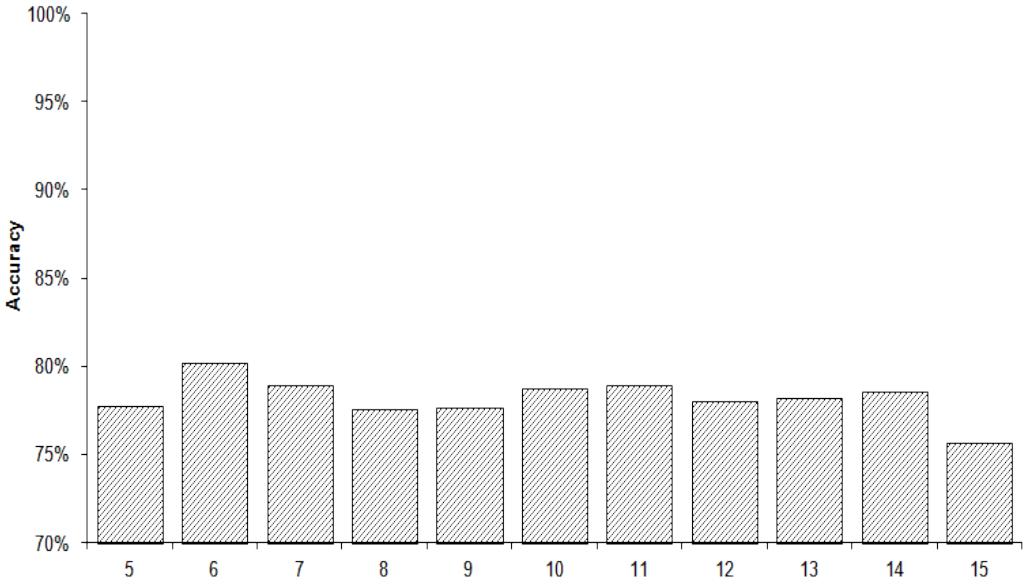


Figure 6.8: HMM number of hidden states N versus classification accuracy.

The highest overall accuracy was achieved when $N = 6$ as shown in Fig. 6.8. Therefore the number of hidden states N was set to $N = 6$ in for experimental validation.

6.3.2 OSW, DTW and HMM with Online and Inaccurate Activity Segmentation

The SW approach is able to locate and quantify optimal subsequences embedded within another sequence, referred to as local sequence alignment in bioinformatics. It is able to do this as the zero value terminates poor subsequence alignments, corresponding to negative similarity scores, and optimally finds and quantifies the maximum corresponding spatial subsequence(s) between two sequences. This local alignment characteristic allows SW to be applied to spatial sequences, without the need for accurate activity sequence segmentation. In the following experiment, threshold-based NN classification is conducted with inaccurately segmented sequences of varying sizes in contrast to DTW and the HMM. The purpose of this experiment is to demonstrate the effectiveness of the efficient OSW approach in real world online applications.

The experiment comprises 12 activities, which are equally separated into training and testing sets, with thresholds empirically derived by measuring the average intraclass distance between training sequences and/or templates ± two standard deviations. With the intraclass thresholds and the derived optimum parameters in section 6.3.1, a threshold-based NN classification experiment is carried out with the online data stream comprising the known activities and using different window sizes. Initially, the window size w is set to the length of the longest activity in the training set. w is then increased by by 5% and 10% of the length of the longest sequence to observe the affect of larger windows sizes on the three approaches. A true positive (TP) occurs when values of the correct activity exceed their specified threshold within the ground truthed online sequence, with no other class template from other activities exceeding corresponding thresholds. A false positive (FP) occurs if any incorrect activity exceeds their threshold within the ground-truthed online sequence. The precision (5.2) and recall (5.3) statistics of the online evaluation are shown in Table 6.3.

Table 6.3: Threshold-based NN classification with Online Recognition.

	w=100%		w=105%		w=110%	
	Precision	Recall	Precision	Recall	Precision	Recall
HMM	70.00%	5.83%	0.00%	0.00%	0.00%	0.00%
DTW	58.62%	42.5%	59.38%	31.67%	58.33%	23.33%
SW	83.05%	81.67%	83.90%	82.50%	82.75%	83.33%

With the window size equal to the length of the longest sequence in the class template set (w=100%), OSW is still able to achieve high precision and recall with the given synthetic online sequence in contrast to the HMM and DTW. Furthermore, these high values were consistent with the larger evaluated windows sizes of $w = 105\%$ and $w = 110\%$ as seen in Table 6.3. The observed high precision and recall of the OSW algorithm across the different window sizes can be explained by the SW technique optimally locating embedded patterns (local alignments) within the online window sequence and terminating poorly matching local alignments arising from significant gaps or mismatches. These poorly matching subsequences generate regions of local negative similarity which are terminated by the zero condition in the relation specified in (6.3).

The results in Table 6.3 demonstrate that DTW is sensitive to extraneous elements from

window sequences and furthermore is sensitive to the specified window size. This can be seen by the reduction in recall with the increase in online window size. As DTW is global and accounts for the additional, non-activity elements of the window sequence the calculated distance typically increases with increasing window size preventing recognition with derived thresholds. Overall DTW does manage to maintain its precision across the evaluated window sizes and does achieve precision and recall values higher than the HMM.

The discrete HMM also was not able to recognise the observed window sequences across the different window sizes. The HMM's high sensitivity to the window size, resulting in a low precision and recall, is due to the log likelihood of $Pr(O|\lambda)$, where O is the observed window sequence and λ is the derived model for an activity, encountering symbols with zero probability, thus resulting in a log likelihood of negative infinity. These zero symbol probabilities occur due to the failure to observe such symbols in the training sequences during HMM parameter estimation. The second and less significant reason for HMM's poor performance is due to the forward inferencing algorithm encountering significant numbers of symbols with low probability in the online window sequence (due to the global matching nature of HMM inferencing). As a result the derived log likelihoods are reduced such that they do not exceed the specified thresholds and the activities are not recognised.

6.3.3 SW, DTW and HMM with Accurate Activity Segmentation

OSW is capable of online activity recognition with inaccurate activity segmentation; however, the discrimination ability of the SW approach with accurately segmented activities is also of importance for applications that can accurately quantify activity start and end points. To evaluate this aspect accurately segmented training sequences from the 12 activity classes of dataset C (2.6) are used and intraclass distances derived. With the applied thresholds for each activity, threshold-based NN classification is used with the accurately segmented test sequences and the results contrasts to DTW and the discrete HMM. The results are shown in Table 6.4.

These results show that the SW approach is capable of producing high precision (98%) and recall (97%) in classification of accurately segmented spatial activity sequences with

Table 6.4: Threshold-based NN classification with Accurate Activity Segmentation.

	Precision (%)	Recall (%)
HMM	83.9	75.6
DTW	97.3	96.9
SW	98.1	97.6

the given data set. Furthermore, the recognition performance of SW is comparable to the global DTW approach. This finding demonstrates that it is possible to apply the local SW alignment technique successfully in a global matching role with innate activity temporal variation. In contrast, both the SW and DTW alignment approaches significantly outperformed the HMM, further providing evidence of the strong discriminatory capability of the proposed SW alignment approaches.

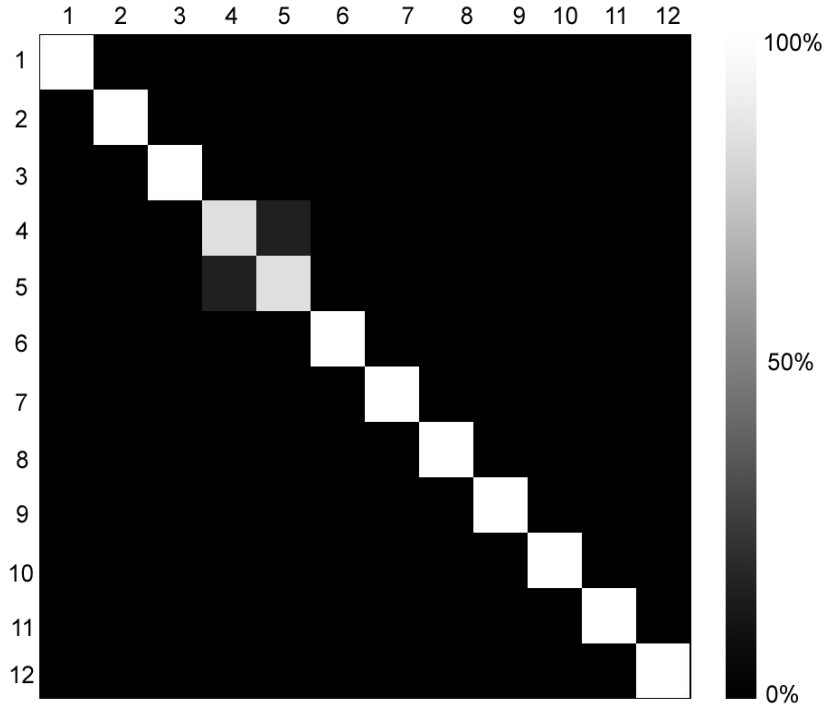


Figure 6.9: Confusion matrix for SW. The legend represents the percentage of activity sequences classified.

Elaborating further on the SW results, the SW confusion matrix, presented in Fig.

6.9, demonstrates near perfect classification with only minor errors occurring between activities four and five, which represent two variations of having breakfast. The misclassification seen here is understandable as the variants of having breakfast share 90% spatial similarity (see Fig. 6.10) and the remainder of the sequences have only a small spatial disparity.

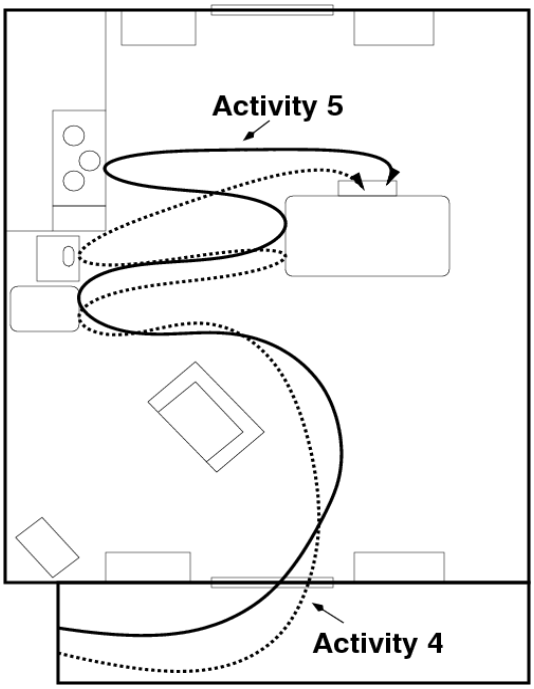


Figure 6.10: Spatial patterns for confused activities 4 and 5.

To draw further insight on which of the three approaches HMM, DTW or SW are more appropriate for the task of online recognition and activity segmentation, one needs to contrast the results from Table 6.4 with the computational complexities of the techniques. If one looks at the discrete HMM described in Section 2.3.2 and applied here, the learning of each discrete HMM model λ is achieved through iterative application of the Baum-Welch algorithm, with the number of iterations I determined through convergence of the ratio of the average of the log-likelihoods between the current and previous iterations to a specified threshold. As described in (Rabiner, 1989), the complexity of the Baum-Welch algorithm is $O(TN^2)$, where N is the number of hidden states and T is the sequence length. To derive the final complexity of learning λ from a set of sequences one simply multiplies the complexity of performing the Baum-Welch algorithm

by the number of iterations required to converge, resulting in $O(ITN^2)$. With λ one can estimate the $Pr(w|\lambda)$ of a window sequence w using just the forward algorithm, which exhibits a complexity of $O(TN^2)$, where $N \ll T$. When HMMs are applied to online sequence segmentation, window sequences are segmented on their periphery only when the resulting $Pr(w|\lambda)$ exceeds a threshold.

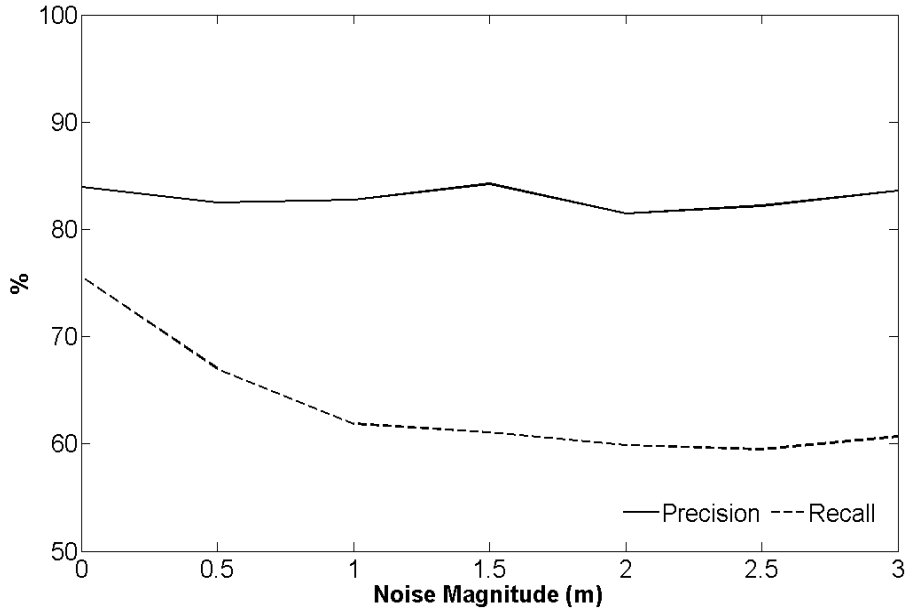
As DTW and SW are template rather than generative approaches they do not require supervised learning like the HMM. However, they are both DP-based and require initialisation of DP matrices, which requires $O(wC_i)$ time and space for a window sequence w and class sequence C_i. In online streaming applications, the two approaches differ in that DTW still requires $O(wC_i)$ time to calculate its full DP matrix at each window, while OSW requires only $O(C_i)$ due to the need to only calculate the last row as shown in algorithm 5.

Similar to the HMM, DTW will segment an activity stream at the window periphery when the resulting DTW distance is less than a specified threshold. OSW is different as it requires a traceback procedure to identify the beginning of a matched subsequence when a similarity threshold is reached, requiring an additional $O(w)$ calculations. It is important to note that DTW and the HMM are limited to only segmenting on the window boundaries and can't segment on activity subsequences within the window like OSW. Considering this, computational complexities and the resulting precision and recall shown in Table 6.4, OSW is the preferred of the three approaches for accurate activity segmentation, particularly if those activities are incomplete or abbreviated compared to the exemplar set.

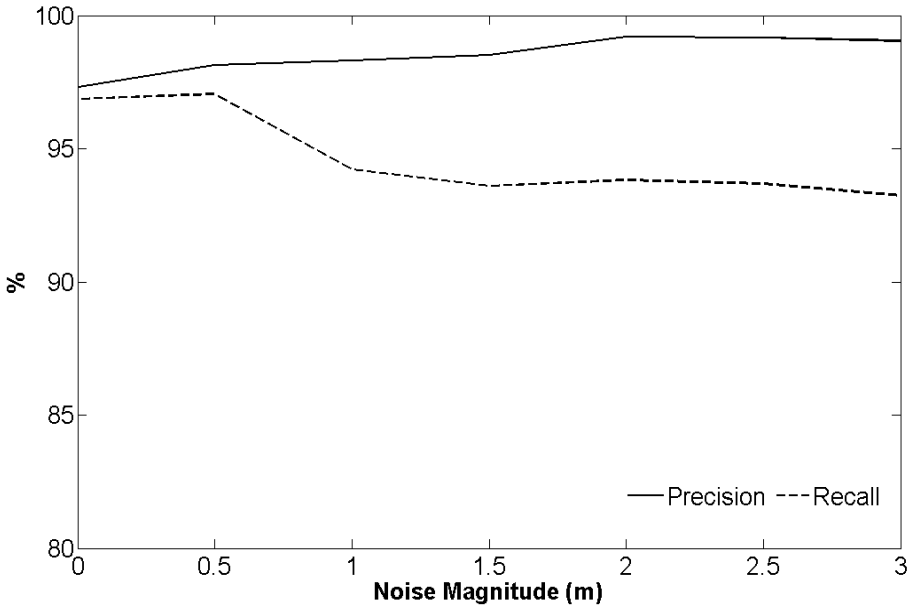
6.3.4 Robustness of SW, DTW and the HMM with Accurately Segmented Activities

Robustness to noise is an important characteristic of any spatial activity recognition approach as video tracking systems typically produce spatial sequences intertwined with noise and gaps. To evaluate the robustness of the modified SW algorithm Gaussian noise is introduced with varying magnitudes into each of the accurately segmented testing sequences during threshold-based NN classification. Benchmarking of the robustness is conducted in relation to the the global DTW technique and the HMM. Experimental

results with noise magnitudes of between 0 to 3 metres are shown in Fig. 6.11.

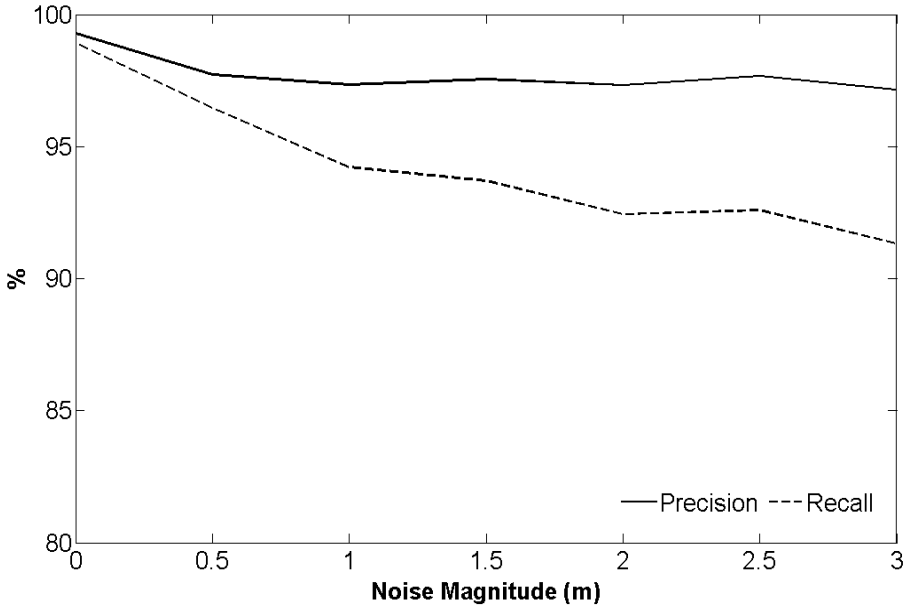


(a) HMM



(b) DTW

The results from the given data set demonstrate that the modified SW algorithm is



(c) SW

Figure 6.11: Noise magnitude versus classification accuracy.

more resilient to noise than the HMM, as indicated by the smaller decrease in recall with the increased magnitudes of noise: 3% decrease for SW versus 14% for the HMM. Importantly, SW was able to maintain a precision and recall of $> 93.0\%$ with the largest evaluated magnitude of noise, while the discrete HMM achieved only 60%. The observed maintenance of high accuracy across the different magnitudes of artificially introduced noise also reinforces the belief that the proposed SW approach is more robust to noise than the discrete HMM.

DTW was also seen to perform similar in relation to noise as the SW approach. This was not expected as DTW is known to be sensitive to noise (Chen *et al.*, 2005) and thus should have exhibited a decrease in recall. It is implicit that introduction of artificially introduced noise increases the average DTW distance to the class templates. As NN classification is used with thresholding, an increase in distance will prevent sequences from being recognised as they are more likely to exceed the specified thresholds. Taking this in to account, it is likely that the derived thresholds, taken from the average intra-class distances of the training sets ± 2 standard deviations, were sufficiently large such that the increased distances did not exceed the thresholds and thus did not affect the

recall statistics.

6.4 Summary

An automatic activity recognition system must capable of recognising human activities in continuous streams of data in real world situations. This chapter explores this problem space and provides a SW local alignment based approach, and a variant (OSW) that is optimised for activity recognition and processing of spatial data streams. The use of a bioinformatics inspired approach in the spatial activity recognition domain is novel. The unique local alignment property of the SW algorithms allow efficient similarity quantification of embedded and partial spatial activities within sliding windows, preventing the need for accurate sequence segmentation. This local alignment (subsequence matching) ability is achieved by the approach penalising and terminating poorly aligned or mismatched sequences, which correspond to embedded or partial activities. The ability to recognise spatial activity sequences from online video tracking systems is significant as traditional techniques focus on applying full window sequences to models or template recognition approaches, and are heavily reliant on accurate stream segmentation for recognition.

The ability of the SW and OSW approaches to recognise embedded activities is demonstrated through the evaluation of a 12 class dataset with online classification. The innate discrimination ability of the SW approach is also evaluated and shown to outperform DTW and HMM techniques with accurately segmented activity sequences. Evaluation of spatial sequences containing various magnitudes of Gaussian noise also confirmed that the SW approach is robust to noise from video tracking systems.

Chapter 7

Conclusion

The use of an automatic activity recognition system in domains such as surveillance, monitoring, and smart homes, allows the identification of human activities and the provision of an automatic response. In a smart home, this is important in order to provide assistance to patients, the elderly and disabled individuals allowing them to maintain independence and reducing support costs.

Activity recognition is a complex problem due to the variable nature with which humans conduct activities. The same activity can have different spatial, temporal, and sequential orders. Furthermore, they can be interweaved amongst other activities at any point. The task of recognition of activities amongst all of this human variation is compounded by the fact that multi-camera video tracking systems generate significant noise, particularly in smart home, multi-room residences, as a result of computer vision challenges. This thesis specifically addresses the problem of recognising human activities with intrinsic spatial and temporal variation in the presence of noise (generated by video tracking systems). This recognition task is achieved through an investigation of biological paradigms and bioinformatics inspired approaches.

Biological and bioinformatics sequence alignment approaches have many useful qualities for pattern recognition including robustness to noise and tolerance to variation (insertions, deletions and substitutions). Several novel approaches are provided in this thesis that exhibit high levels of spatial activity discrimination, yet are tolerant to high levels of noise and innate temporal variation amongst accurately and inaccurately segmented sequences. Another approach is provided that is capable of recognising simple interwoven activities, which is important for recognising real world human activities.

The cellular chemotactic model is provided in Chapter 3 to address the robustness issue with noisy spatial activity sequences. Bacteria are believed to have evolved the

chemotactic capability to sense and respond to dynamic environments, increasing their survivability and thus evolutionary fitness. Through the use of a chemotactic paradigm, the cellular chemotactic model is shown to be robust to noise from video tracking systems with empirically estimated parameters. Furthermore, the biologically inspired model exhibits inherent resilience to spatial variations in activity sequences of similar duration, due to the absence of Markovian constraints for matching. The approach was compared to a discrete HMM, producing superior discrimination results, even with small numbers of templates.

One of the most difficult aspects of activity recognition is the ability to recognise interwoven or multi-tasked activities. This is due to humans interweaving activities at any time between one or more activities. The cellular chemotactic model addresses this multi-tasking recognition issue as to its cells mimic the behaviour of agent-based models. In this instance multiple "activity" cells move towards an "attractant" when subsequences correspond (partial activities), and conduct unbiased random walks in the presence of non-matching subsequences (areas of interweaving or noise).

In Chapters 4 and 5 sequence alignment techniques from bioinformatics are used as a source of inspiration to continue the investigation of robust spatial activity recognition approaches. Sequence alignment techniques are regularly used with biological sequences for quantitative and qualitative similarity comparisons. Biological sequences exhibit similar issues to those found with human activities: compression (relating to shorter duration activities), expansion (relating to longer duration activities), insertions and deletions (relating to activity variability and tracking noise) and partial sequences (relating to activity interweaving). Therefore, sequence alignment approaches are capable of dealing with the challenges of spatial activity recognition. In Chapter 4 fusion of sequence alignment with these bioinformatics "time warping" characteristics results in the formulation of the LCSD and GED approaches for dealing specifically with noise tolerance. An empirical approach is taken to estimate suitable parameters for accurate spatial recognition. These parameters and the respective approaches are then validated against other sequence alignment, time warping and probabilistic approaches. The results demonstrate that LCSD and GED are more robust to noise than DTW with accurately segmented activities and have improved discrimination with spatially similar activities.

Using sequence alignment characteristics, a threshold DTW (TDTW) approach is specif-

ically developed for recognition of activities exhibiting temporal variation, whilst having an improved tolerance to tracking system noise. TDTWs robustness is achieved via the introduction of a novel sequence alignment distant matching constraint in the TDTW calculation for spatial elements. The constraint prevents minor warping with small changes in position, reducing the algorithms susceptibility to tracking noise. The innate temporal and spatial variation of 10 human activities captured from a multi-camera video tracking system is also verified empirically by analysis of the sequence composition and length variability. The TDTW approach is validated using this dataset against time warping and probabilistic algorithms to demonstrate its superior discrimination with spatial activity sequences. The runtime performance of TDTW is further improved via application of a band DP constraint, which results in only minor decreases in recognition performance with smaller band sizes.

Chapter 6 continues with another bioinformatics inspired approach, but focuses on activity recognition from continuous data streams. In traditional activity recognition approaches, data streams are processed using a sliding window approach. Extracted and segmented sequences are then analysed using a pattern recognition technique such as DTW or a HMM. Sliding windows of different sizes need to be provisioned as most activity recognition approaches are unable to detect activities embedded in a larger sequence. The SW local alignment algorithm is adapted in Chapter 6 to a two dimensional spatial activity recognition context. The SW approach allows one to efficiently locate and quantify similarity of embedded spatial activity sequences in a spatial data stream, as well as detect optimal subsequences with only partial activities. The ability to recognise spatial activity sequences from online video tracking systems is significant as traditional approaches focus on applying full window sequences to models or template recognition. The SW approach formulated in this thesis has been further optimised in regards to efficiency for continuous spatial data streams via the OSW approach. Experimental validation with existing sequence alignment and activity recognition approaches confirms the high recognition performance of the developed SW and OSW approaches with both accurate and inaccurately segmented activities.

7.1 Future Directions

This thesis has shown that biological paradigms and bioinformatics can be used as inspiration for addressing pattern recognition problems, in particular spatial activity recognition (constrained to a smart home environment). The biologically inspired cellular chemotactic model presented in Chapter 3 is tolerant to tracking system noise and spatial sequence variability like its biological process. Due to its multi-agent characteristics, the model can identify simple interwoven activities. However, this approach has limitations that affect its ability to be used for real world activity recognition. These limitations and areas of possible future research are given below:

- The model is susceptible to activities exhibiting significant intra-activity temporal variation. One such real world example of this type of activity is watching TV which can have significant differences in duration depending on what a person is watching on the TV and when they are watching it (weekday versus weekend, morning versus evening). One such mechanism of improving the models susceptibility to this type of temporal variation is by modelling the compression and expansion of the relevant spatial sequences as a function of the environments "attractant" gradient and the velocity of the activity cell. In the event of sequence expansion (resulting from an increased duration of an activity) an activity cell will reduce its biased random walk (and hence velocity) towards an attractant source (along a larger gradient) such that further expansion results in smaller cell movements. This process mimics bacterial chemotaxis, whereby bacterial cells reduce their motility in higher concentrations of an attractant and eventually return to non-attractant patterns of behaviour to prevent localisation. The current model only employs a very simplistic abstraction of this concept.

- The cellular chemotactic model is currently reliant on a sliding window approach or accurate activity segmentation for recognition of spatial activities within online data streams. This occurs as the model remains in vicinity of an "attractant" source post-matching, requiring the environment to be reset for further recognition tasks. Simple interwoven activities can be recognised within this limitation, if spatially activities overlap. To allow online recognition without sliding windows or activity segmentation, the chemotactic model could be expanded to a live agent-based approach in a multi-dimensional environment with multiple "attractants".

For this to occur, a mechanism would need to be established to allow the cells to return to a uniformly distributed, homogenous and unbiased random walking state in the environment. To clarify this concept further, activity cells would move towards "attractant" sources when they have matching elements of sequences and after a match is found and signalled, would return to a steady state, prior to commencement of further matching. In this context, the cells would be "live" allowing automatic recognition with minimal intervention. The approach would still require supervised classification of activity sequences to form the respective activity cells.

- Complex activity recognition, whereby a set of simple activities form more complex activities, cannot be represented in this model. An example of a complex activity is *get-ready-for-work*, which can consist of simple activities: *have-shower*, *brush-teeth*, *get-dressed*, and *have-breakfast*. Complex activity recognition can be explored in further research through the use of evolution Engelbrecht (2005), modularity (Alon, 2003; Rives and Galitski, 2003), colony formation (Passino, 2002) or multicellularity (Bonner, 1998) paradigms. These concepts can be readily applied to the cellular chemotactic model to aggregate activity cells that belong to a complex activity to form "modules", "colonies" or "multicellular organisms". The process of evolution can then be applied to select those aggregates that best represent the complex activity.

- One of the key limitations of the cellular chemotactic model is is lack of specificity, resulting from the receptor and activity sequence abstraction. In biological systems, receptors are complex three dimensional structures that exhibit dependencies between other receptors, and can be influenced by cellular and environmental factors. In retrospect, an approximate receptor binding (lock-and-key) model would be better suited for the spatial activity domain. Using this concept, subsequences with high degrees of similarity would exhibit better receptor matching resulting in increased motility. On the other hand, subsequences with poor matching would exhibit weaker receptor matching and thus decreased motility. Like most biological systems, receptors are subject to evolutionary pressures. This concept could be applied to the chemotactic model to optimise receptor binding over time via selectively modifying receptors and receptor compositions, and then selecting only those cells with improved recognition. Mutation rates for sequences could be empirically ascertained by analysing the variability amongst spatial sequences of the same activity. This mutation and optimisation concept is valuable for recognition

of human activities due to their dynamic nature. Evolutionary mechanisms from biology can be explored in further research to allow exchange of subsequence information. Genetic transfer and cross-over of biological information is common in nature and is seen to provide diversity and resilience in dynamic environments. Applying such a mechanism to activity sequences can be used to generate new activities or more optimal templates of activities. These activities exhibiting cross-over mutations could in turn be subject to fitness functions and thus be selected to provide resilience to changes in human activities.

In Chapters 4 - 6 of this thesis, sequence alignment and time warping approaches were explored in the spatial activity recognition context. The LCSD, GED, TDTW and SW approaches were all demonstrated to provide good activity discrimination and robustness to noise in a smart home environment. The following outlines limitations of the approaches and recommendations for future research:

- Parameter estimation. A limitation of all of these approaches is that the parameters are derived via empirical experiments with training or template sequences across a range of parameter values. In the case of the SW approach, its parameters are less sensitive in regards to discrimination performance; however, optimisation could still be useful. To provide a less empirical approach to determining optimal parameter values, analysis of the template or training data is required to quantify the variability between sequences (resulting from gaps due to insertions, deletions or substitutions). This information could be derived via a multiple sequence alignment approach as outlined in Waterman *et al.* (1991) or McClure *et al.* (1994), or through probabilistic means such as used in Wei (2004). An analytical approach will provide insight as to the variability of the spatial data to assist with parameter estimation. The intent is to use this analytical data on the spatial sequence variability to determine a relationship between parameter values and variability, and then derive a heuristic to estimate optimal parameters.
- Alternate gap and mismatch models. A linear gap model was found in GED and SW approaches to achieve the most consistent recognition performance across the respective datasets. This requires further investigation in consultation with the analytical parameter estimation approach, as emplar-based analytics could provide insight on the gap variability and sequence mismatches in a group of activities. A probabilistic gap and mismatch model will likely capture this variability more

effectively resulting in a gap and mismatch model that better approximates the variability of an activity, and further discriminates spatially distinct exemplars.

- Hybrid recognition approaches. The sequence alignment approaches that were investigated performed well with accurately segmented activity sequences in the presence of tracking system noise. The LCSS and SW similarity-based approaches performed extremely well in the presence of significant noise and minor activity variability as they limited the quantification of noise in their similarity calculation. The LCSD, GED and TDTW distance-based approaches also performed well discriminating accurately segmented activities, but performed better with discrimination of spatially similar activities. A hybrid approach consisting of both similarity and distance-based techniques could thus be explored to provide the noise abating characteristics of LCSS and SW, whilst achieving high discrimination with spatial and temporally similar activities as per LCSD, GED and TDTW.

It is acknowledged that recognition of spatial activities from video-based tracking systems is challenging. The SW and OSW approaches presented in Chapter 6 provides a solid mechanism for dealing with inaccurately segmented activity sequences in the presence of noise and minor activity variation. The OSW approach provided in this thesis outlines a more efficient formulation that specifically deals with the problem of online recognition. As OSW can identify approximate start and end points of sequences via traceback procedures and optimal similarity values, OSW could be applied for segmentation of online data streams.

Finally, the incorporation of environmental sensor data, collected from other sensors in a smart home such as touch, contact or thermal sensors, could assist with dynamically improving the performance of these spatial activity approaches via adjusting parameters in relation to sensor feedback in a control systems approach.

Bibliography

Aach, J. and Church, G. (2001). Aligning gene expression time series with time warping algorithms. *Bioinformatics*, **17**(6), 495–508.

Adler, J. (1975). Chemotaxis in bacteria. *Annu. Rev. Biochem.*, **44**, 341–356.

Aggarwal, J. and Cai, Q. (1999). Human motion analysis: a review. *Computer Vision and Image Understanding*, **73**(3), 428–440.

Aggarwal, J. and Park, S. (2004). Human motion: modeling and recognition of actions and interactions. In *IEEE Int. Conf. on 3D Data Processing, Visualization and Transmission (2DPVT04)*, pages 640–647.

Aggarwal, J. and Ryoo, M. (2011). Human activity analysis: A review. *ACM Computing Surveys*, **43**(3), 16:1–16:43.

Alon, U. (2003). Biological networks: The tinkerer as an engineer. *Science*, **301**(5641), 1866–1867.

Alon, U., Surette, M., Barkai, N., and Leibler, S. (1999). Robustness in bacterial chemotaxis. *Letters to Nature*, **397**, 913–917.

Andrade, E., Blunsden, S., and Fisher, R. (2006). Modelling crowd scenes for event detection. In *IEEE Int. Conf. on Pattern Recognition (ICPR06)*, pages 175–178.

Apostolico, A. and Giancarlo, R. (1998). Sequence alignment in molecular biology. *Journal of Computational Biology*, **5**(2), 173–196.

Barkai, N. and Leibler, S. (1997). Robustness in simple biochemical networks. *Nature*, **387**(6636), 913–917.

Barton, G. (1998). Protein sequence alignment techniques. *Acta Crystallographica Section D: Biological Crystallography*, **54**(6), 1139–1146.

Ben-Arie, J., Wang, Z., Pandit, P., and Rajaram, S. (2002). Human activity recognition using multidimensional indexing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **24**(8), 1091–1104.

Berg, H. (1990). Bacterial microprocessing. In *Cold Spring Harbor Symposia on Quantitative Biology*, volume 55, pages 539–545.

Bhalla, U. (2003). Understanding complex signaling networks through models and metaphors. *Progress in Biophysics & Molecular Biology*, **81**(1), 45–65.

Bobick, A. and Ivanov, Y. (1998). Action recognition using probabilistic parsing. In *IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, pages 196–202.

Bobick, A. and Ivanov, Y. (2001). The recognition of human movement using temporal templates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **23**(3), 257–267.

Bonner, J. (1998). The origins of multicellularity. *Integrative Biology Issues News and Reviews*, **1**(1), 27–36.

Bourret, R. and Stocks, A. (2002). Molecular information processing:lessons from bacterial chemotaxis. *The Journal of Biological Chemistry*, **277**(12), 9625–9628.

Bui, H., Venkatesh, S., and West, G. (2001). Tracking and surveillance in wide-area spatial environments using the abstract hidden markov model. *Int. Journal on Pattern Recognition and Artificial Intelligence*, **15**(1), 177–195.

Chan, M., Camp, E., Esteve, D., and Fourniols, J. (2009). Smart homes - current features and future perspectives. *Maturitas*, **64**, 90–97.

Chen, L. and Ng, R. (2004). On the marriage of lp-norms and edit distance. In *30th VLDB Conf.*, pages 792–803.

Chen, L., Özsu, M. T., and Oria, V. (2004). Symbolic representation and retrieval of moving object trajectories. In *6th ACM SIGMM Int. Workshop on Multimedia Information Retrieval*, pages 227–234.

Chen, L., Özsu, M. T., and Oria, V. (2005). Robust and fast similarity search for moving object trajectories. In *24th ACM Int. Conf. on Management of Data*.

Chu, S., Keogh, E., Hart, D., and Pazzani, M. (2002). Iterative deepening dynamic time warping for time series. In *SIAM Int. Conf. on Data Mining*, pages 195–212.

Crochemore, M. and Rytter, W. (2002). *Jewels of Stringology*. World Scientific Press.

Das, S. (1982). Some experiments in discrete utterance recognition. *IEEE Trans. on Acoustics, Speech and Signal Processing*, **ASPP-30**(5), 766–770.

de Castro, L. and Timmis, J. (2002). *Artificial immune systems: a new computational intelligence approach*. Springer-Verlag, London.

Dilger, W. (1997). Decentralized autonomous organization of the intelligent home according to the principle of the immune system. *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*, **1**, 351–356.

Dorigo, M., Caro, G. D., and Gambardella, L. (1999). Ant algorithms for discrete optimization. *Artificial life*, **5**(2), 137–172.

Duong, T., Bui, H., Phung, D., and Venkatesh, S. (2005). Activity recognition and abnormality detection with the switching hidden semi-markov model. In *IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 838–845.

Ehlert, P. (2003). Intelligent user interfaces: introduction and survey. Technical Report DKS03-01 / ICE 01, Delft University of Technology, Netherlands.

Eidhammer, I., Jonassen, I., and Taylor, W. (2004). *Protein bioinformatics:an algorithmic approach to sequence and structure analysis*, chapter 1.2, pages 3–23. John Wiley & Sons Ltd.

Engelbrecht, A. (2005). *Fundamentals of Computational Swarm Intelligence*, chapter 2, pages 55–84. John Wiley & Sons Ltd.

Fang, H., Srinivasan, R., and Cook, D. (2012). Feature selections for human activity recognition in smart home environments. *International Journal of Innovative Computing, Information and Control*, **8**(5b), 3525–3535.

Fuentes, L. and Velastin, S. (2006). People tracking in surveillance applications. *Image and Vision Computing*, **24**(11), 1165–1171.

Gotoh, O. (1982). An improved algorithm for matching biological sequences. *Journal of Molecular Biology*, **162**, 705–708.

Guo, A. and Siegelmann, H. (2004). Time-warped longest common subsequence algorithm for music retrieval. In *Int. Conf. on Music Information Retrieval*.

Hagras, H., Colley, M., Callaghan, V., Clarke, G., Duman, H., and Holmes, A. (2002). A fuzzy incremental synchronous learning technique for embedded-agents learning and

control in intelligent inhabited environments. In *IEEE Int. Conf. on Fuzzy systems*, pages 139–145.

Hamid, R., Huang, Y., and Essa, I. (2003). Argmode - activity recognition using graphical dodels. In *IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, volume 4, pages 38–44.

Hauptmann, A., Gao, J., Yan, R., Qi, Y., Yang, J., and Wactlar, H. (2004). Automated analysis of nursing home observations. *Pervasive Computing*, **3**(2), 15–51.

Haykin, S. (1999). *Neural Networks: A comprehensive foundation*. Prentice Hall, New York, USA, 2nd edition.

Hine, N., Judson, A., Ashraf, S., Arnott, J., Sixsmith, A., Brown, S., and Garner, P. (2005). Modelling the behaviour of elderly people as a means of monitoring well being. *Lecture Notes in Computer Science*, **3538**, 241–250.

Intille, S. and Bobick, A. (1998). Representation and visual recognition of complex, multi-agent actions using belief networks. Technical report, M.I.T Media Laboratory Perceptual Computing Section. No. 454.

Itakura, F. (1975). Minimum prediction residual principle applied to speech recognition. *IEEE. Trans. on Acoustics, Speech and Signal Processing*, **ASSP-23**(1), 67–72.

Ivanov, Y. and Bobick, A. (2000a). Recognition of visual activities and interactions by stochastic parsing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(8), 852–872.

Ivanov, Y. and Bobick, A. (2000b). Recognition of visual activities and interactions by stochastic parsing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(8), 852–872.

Kahveci, T. and Singh, A. (2001). Variable length queries for time series data. In *17th Int. Conf. on Data Engineering*, pages 273–282.

Keogh, E. and Pazzani, M. (2001). Derivative dynamic time warping. In *SIAM Int. Conf. on Data Mining*, volume 1, pages 5–7.

Kim, E., Helal, S., and Cook, D. (2010). Human activity recognition and pattern discovery. *IEEE Pervasive Computing*, **9**(1), 48–53.

Kirkpatrick, S., Jr., C. G., and Vecchi, M. (1983). Optimization by simulated annealing. *Science*, **220**(4598), 671–680.

Lee, S. and Mase, K. (2002). Activity and location recognition using wearable sensors. *Pervasive Computing*, **1**(3), 24–32.

Lühr, S., Bui, H., Venkatesh, S., and West, G. (2003). Recognition of human activity through hierarchical stochastic learning. In *IEEE Int. Conf. on Pervasive Computing and Communications (PerCom-03)*, pages 416–422.

MacNab, R. and Koshland, D. (1972). The gradient-sensing mechanism in bacterial chemotaxis. *Proc. Nat. Acad. Sci.*, **69**(9), 2509–2512.

McClure, M., Vasi, T., and Fitch, W. (1994). Comparative analysis of multiple protein-sequence alignment methods. *Molecular Biology and Evolution*, **11**(4), 571–592.

Moeslund, T., Hilton, A., and Kruger, V. (2006). A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, **104**(2), 90–126.

Monk, T. H., Reynolds, C. F., Machen, M. A., and Kupfer, D. J. (1992). Daily social rhythms in the elderly and their relation to objectively recorded sleep. *Sleep*, **15**(4), 322–329.

Needleman, S. B. and Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, **48**, 443–453.

Nguyen, N., Venkatesh, S., West, G., and Bui, H. (2002). Coordination of multiple cameras to track multiple people. In *Asian Conference on Computer Vision*, pages 302–307.

Nguyen, N., Bui, H., Venkatesh, S., and West, G. (2003). Recognising and monitoring high-level behaviours in complex spatial environments. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR-03)*.

Oliver, N., Rosario, B., and Pentland, A. (2000). A bayesian computer vision system for modeling human interactions. *IEEE Transactions on Pattern Analaysis and Machine Intelligence*, **22**(8), 831–843.

Passino, K. (2002). Biomimicry of bacterial foraging for distributed optimization and control. *IEEE Control Systems*, **22**(3), 52–67.

Paton, R., editor (1994). *Computing with biological metaphors.* Chapman & Hall, UK.

Peursum, P., Venkatesh, S., West, G., and Bui, H. (2003). Object labelling from human action recognition. In *IEEE Int. Conf. on Pervassive Computing and Communications (PERCOM05)*, pages 399–406.

Peursum, P., Bui, H., Venkatesh, S., and West, G. (2004). Human action segmentation via controlled use of missing data in hmms. In *IAPR Int. Conf. on Pattern Recognition*, pages 440–445.

Prlic, A., Domingues, F., and Sippl, M. (2000). Structure-derived substitution matrices for alignment of distantly related sequences. *Protein Engineering*, **13**(8), 545–550.

Pynadath, D. and Wellman, M. (2000). Probabilistic state-dependent grammars for plan recognition. In *16th Ann. Conf. on Uncertainty in Artificial Intelligence*, pages 507–514.

Rabiner, L., Rosenberg, A., and Levinson, S. (1978). Considerations in dynamic time warping algorithms for discrete word recognition. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, **26**(6), 575–582.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, **77**(2), 257–286.

Ratanamahatana, C. and Keogh, E. (2004a). Everything you know about dynamic time warping is wrong. In *Third Workshop on Mining Temporal and Sequential Data.*

Ratanamahatana, C. and Keogh, E. (2004b). Making time-series classification more accurate using learned constraints. In *SIAM Int. Conf. on Data Mining*, pages 11–22.

Rath, T. and Manmatha, R. (2003). Word image matching using dynamic time warping. In *Computer Vision and Pattern Recognition*, pages 521–527.

Rivera-Illingworth, F., Callaghan, V., and Hagras, H. (2005). A neural network agent based approach to activity recognition in ami environments. In *The IEE Int. Workshop on Intelligent Environments*, pages 92–99.

Rives, A. and Galitski, T. (2003). Modular organization of cellular networks. *Proceedings of the National Academy of Sciences*, **100**(3), 1128–1133.

Robertson, N. and Reid, I. (2006). A general method for human activity recognition in video. *Computer Vision and Image Understanding*, **104**(2), 232–248.

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

Sankoff, D. and Kruskal, J., editors (1999a). *Time Warps, String Edits, and Macromolecules:The Theory and Practice of Sequence Comparison*. CSLI Publications, USA.

Sankoff, D. and Kruskal, J., editors (1999b). *Time Warps, String Edits, and Macromolecules:The Theory and Practice of Sequence Comparison*, chapter How to Compute String-Edit Distances Quickly. CSLI Publications, USA.

Sankoff, D. and Kruskal, J., editors (1999c). *Time Warps, String Edits, and Macromolecules:The Theory and Practice of Sequence Comparison*, chapter An overview of Sequence Comparison. CSLI Publications, USA.

Segall, J., Block, S., and Berg, H. (1986). Temporal comparisons in bacterial chemotaxis. *Proc. Natl. Acad. Sci.*, **83**, 8987–8991.

Shah, M. and Jain, R., editors (1997). *Motion-Based Recognition*, chapter Visual Recognition of Activities, Gestures, Facial Expressions and Speech: An Introduction and Perspective. Kluwer Academic Publishers, USA.

Sheikh, Y., Sheikh, M., and Shah, M. (2005). Exploring the space of human action. In *IEEE Int. Conf. on Computer Vision (ICCV05)*, volume 1, pages 144–149.

Smith, T. and Waterman, M. (1981). Identification of common molecular subsequences. *Journal of Molecular Biology*, **147**, 195–197.

Suzuki, R., Ogawa, M., Tobimatsu, Y., and Iwaya, T. (2001). Time-course action analysis of daily life investigations in the welfare techno house in mizusawa. *Telemedicine Journal and e-Health*, **7**(3), 249–259.

Suzuki, R., Ogawa, M., Otake, S., Izutsu, T., Tobimatsu, Y., Izumi, S., and Iwaya, T. (2004). Analysis of activities of daily living in elderly people living alone: Single-subject feasibility study. *Telemedicine Journal and e-Health*, **10**(2), 260–276.

Suzuki, R., Otake, S., Izutsu, T., Yoshida, M., and Iwaya, T. (2006). Monitoring daily living activities of elderly people in a nursing home using an infrared motion-detection system. *Telemedicine Journal and e-Health*, **12**(2), 146–155.

Szalai, A. (1972). *The use of time : Daily activities of urban and suburban populations in twelve countries*. The Hague:Mouton.

Tan, H. and Silva, L. D. (2003). Human activity recognition by head movement using elman network and neuro-markovian hybrids. In *Image and Vision Computing New Zealand*, pages 320–326.

Tapia, E. (2003). Activity recognition in the home setting using simple and ubiquitous sensors. Master's thesis, School of Architecture and Planning, Massachusetts Institute of Technology.

Truyen, T., Bui, H., and Venkatesh, S. (2005). Human activity learning and segmentation using partially hidden discriminative models. In *International Workshop on Human Activity Recognition and Modelling (HAREM2005)*, pages 87–95.

Turaga, P., Chellappa, R., Subrahmanian, V., and Udrea, O. (2008). Machine recognition of human activities: a survey. *IEEE Transactions on Circuits and Systems for Video Technology*, **18**(11), 1473–1488.

Vaswani, N., Chowdhury, A., and Chellappa, R. (2003). Activity recognition using the dynamics of the configuration of interacting objects. In *Computer Vision and Pattern Recognition (CVPR03)*, volume 2, pages 633–640.

Vlachos, M., Kollios, G., and Gunopulos, D. (2002a). Discovering similar multidimensional trajectories. In *18th Int. Conf. on Data Engineering*, pages 673–684.

Vlachos, M., Gunopulos, D., and Kollios, G. (2002b). Robust similarity measures for mobile object trajectories. In *13th Int. Workshop on Database and Expert Systems Applications*, pages 721–726.

Vlachos, M., Gunopulos, D., and Das, G. (2002c). Rotation invariant distance measures for trajectories. In *10th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pages 707–712.

Waterman, M., Joyce, J., and Eggert, M. (1991). Computer alignment of sequences. *Phylogenetic analysis of DNA sequences*, pages 59–72.

Waterman, M. S. (1995). *Introduction to Computational Biology.* Chapman & Hall, London, UK, first edition.

Wei, J. (2004). Markov edit distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **26**(3), 311–321.

Wojek, C., Nickel, K., and Stiefelhagen, R. (2006). Activity recognition and room-level tracking in an office environment. In *IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems*, pages 25–30. IEEE.

Yamato, J., Ohya, J., and Ishii, K. (1992). Recognizing human action in time-sequential images using hidden markov model. In *Computer Vision and Pattern Recognition*, pages 379–385.

Zouba, N., Bremond, F., Thonnat, M., and Vu, V. (2007). Multi-sensors analysis for everyday activity monitoring. In *IEEE Int. Conf on Sciences of Electronic Technologies of Information and Telecommunications (SETIT07)*, pages 25–29.