A hybrid model for capturing implicit spatial knowledge

Corina Sas
Computing Department, Lancaster University
Lancaster, LA1 4WA, UK
c.sas@lancaster.ac.uk

ABSTRACT

This paper proposes a machine learning-based approach for capturing rules embedded in users' movement paths while navigating in Virtual Environments (VEs). It is argued that this methodology and the set of navigational rules which it provides should be regarded as a starting point for designing adaptive VEs able to provide navigation support. This is a major contribution of this work, given that the up-to-date adaptivity for navigable VEs has been primarily delivered through the manipulation of navigational cues with little reference to the user model of navigation.

Categories and Subject Descriptors

H.5 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*Artificial, augmented, and virtual realities* ; I.2.6 [**Artificial Intelligence**]: Learning—*Connectionism and neural nets, Knowledge acquisition*

General Terms

Human Factors

Keywords

user modelling, connectionism, implicit knowledge elicitation, virtual reality

1. INTRODUCTION

This work focuses on understanding how people explore an indoor virtual space. It argues that different navigational patterns, which reflect a user mental model of navigation and are embedded in movement paths, can be actually captured. Attempts to validate this hypothesis require a novel methodology. Traditional techniques for knowledge elicitation present a series of limitations, particularly when it comes to extract implicit knowledge, inherently associated with navigational rules or strategies. Because of their sensitivity to learning temporal sequences, connectionist models, and in particular Recurrent Neural Networks (RNNs) are particularly suitable for extracting such rules [4, 3, 5]. There are three important aspects which support the connectionist approach to modelling navigation. Firstly, navigation is a spatio-temporal process and for this, the particular ability of RNNs to learn temporal sequences represents a major advantage. Secondly, extracting knowledge from the trained RNNs, which have learned to predict the users' trajectory, allows exploring the regularities, implicitly embedded in the trajectory paths. Such regularities can be expressed in terms of rules governing spatial behaviour [9]. Finally, implicitly capturing the navigational rules embedded in movement paths is an unobtrusive process which involves the analysis of user's behaviour rather than user's introspection. Besides its increased objectivity, such a methodology has a significant potential in being automatically used in real-time applications. This is a promising venue for delivering adaptive VEs for navigation support.

Given the transfer of skills from real to the virtual world [11], this investigation can additionally enrich the understanding of human spatial behaviour in the physical world. Apart from the theoretical contributions which such an understanding can provide, it can be also harnessed within practical applications. Designing flexible VEs, able to adapt themselves in order to support user navigation is one of the most promising application fields.

The paper is organised as follows. The next section briefly presents the study design, while the subsequent one provides a detailed presentation of the proposed methodology for capturing the relevant aspects of the mental model of navigation. This led to a set of navigational rules and strategies, underpinning a spatial grammar. Discussion section explores the potential of this work, in terms of linking the user model with the system's potential of adaptivity for supporting user's online behaviour. Other benefits and challenges of the proposed methodology are outlined as well.

2. STUDY DESIGN

The experiment has been carried out within a desktop VE which due to its tractable characteristics permitted the recording of users' positions and headings at each moment in time. Adopting a physical world metaphor the VE consists of a virtual multi-story building where each one of the levels contains three rooms. The rooms have adjacent walls and are connected through doors, offering an intuitive navigational model. The VE has a rectangular shape of 16×27 virtual metres and in order to acquire a complete view, the user has to move and rotate. However, once the user is in a particular room, it usually requires less effort to explore it fully. Figure 1 and Figure 2 offer a bird's eye view of the ground and first floor respectively.

Users can navigate in terms of moving forwards, backwards or rotating, through the use of directional keys. Every time the user presses the up-arrow or down-arrow keys, he/she performs a forward or backward translation. The longer the keys are pressed, the

TAMODIA'05, September 26–27, 2005, Gdansk, Poland.

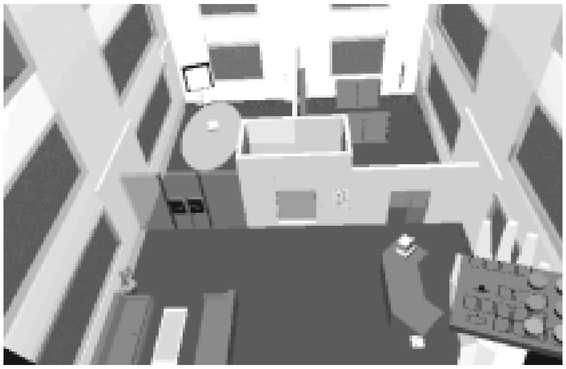

Figure 1: The Bird's Eye View of the Ground Floor

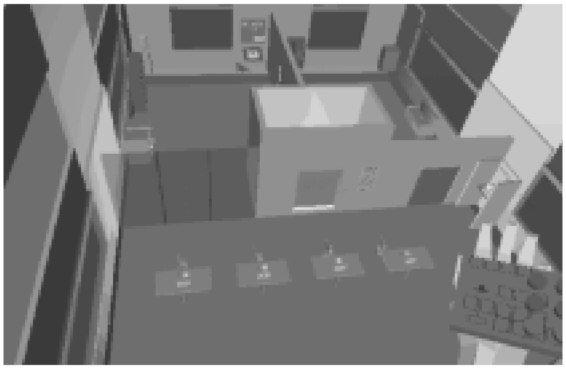

Figure 2: The Bird's Eye View of the First Floor

longer the distance covered within the VEs. Thus, the user moves in a discrete mode, at a constant speed. The height of the view point is the standard height of the avatar, (e.g. 1.70 virtual metre), while the viewing angle through which the user was enabled to perceive the virtual world was 70^o. Users merely use the mouse for selecting a new floor on the panel located in the virtual lift.

The VE system does not provide a predefined set of paths, such as halls or corridors which would limit the user's choice of movements. Therefore, the user can move freely, being restrained only by the walls and objects located on the spatial layout. Since the purpose was to investigate how people explore, search, and acquire spatial information about an indoor environment, this feature of the system has been particularly exploited.

The sample consisted of 32 students: 19 males and 13 females, with an average of 12 years experience of playing computer games. The study involved three phases: familiarisation, exploration and performance measurement. Initially, users were allowed to become accustomed with the VE and to learn movement control. After this, they were asked to perform an exploration task. The exploration task within the virtual building lasted for approximately 25 minutes. After the completion of this task, during which participants acquired spatial knowledge related to the VE, they were tested. Users were placed on the third level and asked to find a particular room located on the ground floor of the virtual building. The time needed to accomplish this task acted as an indicator of the level of spatial knowledge acquired within the VE: the shorter the search time, the better the spatial knowledge.

According to the time required for the search task, users have been identified as *low spatial users*, when they needed significantly longer time to find the library (Mean = 49 seconds), or *high spatial users* who found the library straight away (Mean = 7 seconds).

3. RULE EXTRACTION

This section proposes a hybrid model – symbolic-connectionist – developed for investigating the user mental model of navigation. Given its specific features (feedback which embodies short-term memory) Elman RNNs represent a promising approach of modelling user movement paths. Therefore, an Elman simple RNN [4] was used to learn the trajectory and to predict the next step. Several architectures have been tried in order to minimise the network prediction error. The network prediction has been computed as the sum of squared errors between target and obtained values (SSE), and the mean squared error (MSE) which is an aggregation of the error in the activation levels of the output neurons [14]:

$$MSE = SSE/(NumberVectors - NumberWeights) \quad (1)$$

The architecture which led to the best performance has been retained. It consists of 7 input nodes, 7 hidden nodes, 7 context nodes and 7 output nodes.

The network input consists of a sequence of users' trajectories. At each time step t, an input vector is presented consisting of user's position, orientation angle, distance to the nearest landmark, together with its associated position (coordinates of the centre of the landmark), and the floor where that movement took place. After each trajectory was entered, an input representing "reset" is presented, for which the network is supposed to zero out the outputs [4]. The output pattern represents the input vector of time $t + 1$. Using the backpropagation learning procedure, the network was taught to predict for each current position the next position in time.

The entire set of data was randomly divided into five parts, using two of them for training, one for validation and two for testing. The network was trained with 89 trajectories composed of 13062 input vectors. It was tested with 75 trajectories consisting of 11540 input vectors. The average trajectory length was 160 vectors. The learning rate was 0.001, the initial weights were within the range of (-0.5, 0.5), and the momentum was 0. The network was trained for 1000 epochs and the performances are summarised in Table 1.

Table 1: Summary of Performance Obtained by the RNN Used for Trajectory Prediction

	SSE	MSE	SSE/o-units
Train	1415.33	0.10836	202.191
Test	381.32	0.10845	54.475

The imprecision of floating point arithmetic led to the presumption that a prediction is correct not only if it equals the expected value, but also if it is "close enough" to it. This assumption has high face validity. Therefore in this study it is considered that the RNN produces an error if the Euclidean distance between the vector predicted by the network and the target vector is above a given threshold. This threshold was set up for each element of the vector as follows: 1.5 virtual metre for the (x, y) coordinates of user's position and for the coordinates of landmark's position, 30 degrees for rotation angle, 1 virtual metre for the estimation of the distance to the nearest landmark and 0.3 virtual metre for the estimation of the z coordinate which is related to the prediction of the current floor of the virtual building. Each one of the input elements is predicted with accuracy higher than 78%. The prediction performance obtained by the RNN supports the idea that the net successfully learned the regularities underlying the training data. Understanding what the network learned can be achieved by analysing the internal representation acquired by the network or by a pattern error analysis [8].

Before starting to analyse the pattern error for each of the RNN's predictions, a decision should be made regarding the sample of these predictions which are worth being thoroughly investigated. It is clear that not all the predictions could act as indicators of the regularities embedded in the movement paths, but only the best predictions can qualify for this. Thus, a conservative criterion of selection has been chosen, which has been met only by the best predictions.

The *best predictions* have been identified on the basis of the following performance criterion, which requires that the threshold values should be set up rather low. For a particular input vector: (ix, iy, irot, idist, ixlandmark, iylandmark, ilevel), the predicted user's position or landmark position coordinates could differ only by ± 0.5 virtual metre from the input position coordinates, the predicted user's heading should be higher or smaller with no more than 15 degrees from the input heading, the predicted distance to the nearest landmark could differ only with ± 1 virtual metre from the input distance and the predicted floor value should not differ with more than ± 0.3 virtual metre from the input floor value. These criteria have been met by the top 10% predictions, i.e. the best ones. The next step focuses on analysing how these best predictions occurred, or how the RNN succeeded to predict accurately particular patterns. This issue was addressed through clustering RNN best predictions.

3.1 Clustering Neural Network Predictions

The analysis of the individual pattern error in the network prediction could prove beneficial in understanding the internal representation acquired by the network [8]. The rule extraction process aims to reveal the regularities which allow the RNN to make highly accurate predictions of user's position, heading, nearest landmark and floor.

Given the specifics of these data and the objective of this work a data mining technique has been employed. Self Organising Map (SOM) [7] is based on an unsupervised learning process, allowing cluster identification and visualisation within the input data. This process is carried out without any prior knowledge regarding the number and content of the clusters to be obtained [6].

3.1.1 Pre-Processing Data

Navigation is a spatio-temporal event, where the position of each moment t depends on the position of moments $t-1, t-2, \ldots, t-n$, and in the same time, influences the position at subsequent times: $t+1$, $t+2$, $\ldots$, $t+n$. Thus, one should consider not only the pattern which has been successfully predicted, but also the history in terms of previous patterns which had led to that highly accurate prediction. In other words, the attention is paid to the interesting pattern, considered in its context. Therefore, once a particular pattern has been identified as described above, the context in terms of its previous nine patterns has been also recorded.

For each of these patterns, the predicted values for user's position (px[9], py[9]) and heading (prot[9]) have been retained and concatenated with the corresponding values from the previous patterns. The values have been normalised between -1 and 1. The obtained vector consisted of 30 input elements, three for each of the ten moments in time, and looks like it follows:

$$(px[9], py[9], prot[9], \ldots, px[0], py[0], prot[0]) \quad (2)$$

3.1.2 Training SOM

The training set and the testing set have been identified by analysing the prediction errors associated with the counterpart sets used for training and testing the RNN. The training set consisted of 1367 vectors (54%), while the testing set consisted of 1167 vectors (46%). Each of these two sets covered the top 10% best predictions produced by the RNN, during training and testing respectively.

A SOM of 20×16 neurons was used to perform a topology-preserving mapping. The training parameters were retained after trying more than 100 networks with different architectures and learning rates, because they led to the smallest quantisation error [7] for the testing set: 0.36, while for training set it was 0.31.

3.1.3 Map Visualisation

Training the SOM led to seven clusters of RNN best predictions, as shown in Figure 3. For their identification, within the area corresponding to each of them, the assigned cluster number has been placed. For example, Cluster 1 consists of segments of trajectories standing for best predictions, within area designated by number 1, located down on the right hand side of the map.

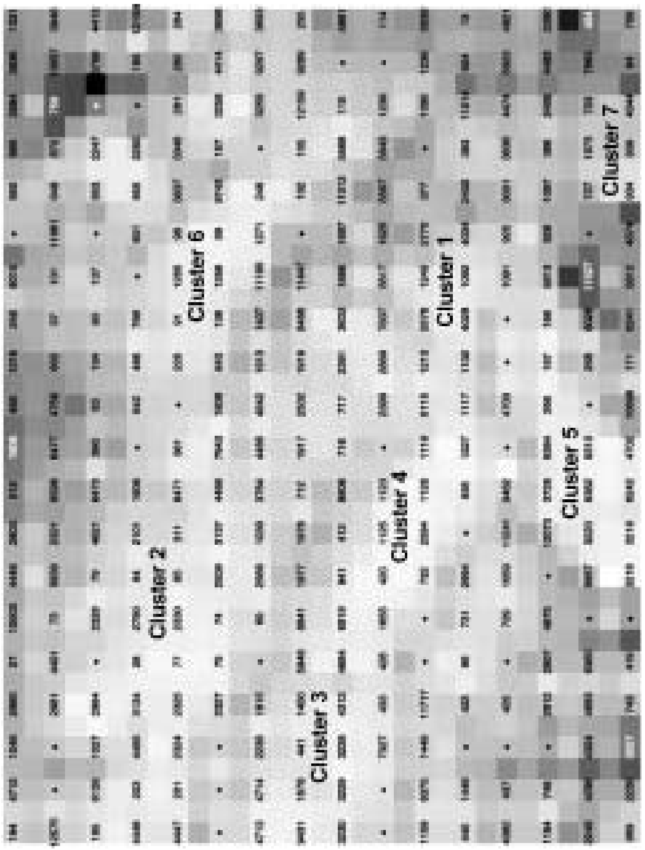

Figure 3: SOM of the Best Predictions of RNN Used for Trajectory Prediction

3.2 Rule Interpretation

The quality of the rules, such as good or poor, determined accordingly to the percent of trajectories performed by efficient navigators vs. inefficient navigators within each cluster, carries an important role in the following descriptions.

There is a significant difference ($\chi^2(6) = 36.96$, $p < 0.01$) regarding the number of best predictions for trajectories performed by high spatial users, comparing with the ones performed by low spatial users, within the previously identified clusters. However, chi-square is an overall test which indicates significant differences if there are significant differences between at least two cell frequencies. In order to identify exactly which pairs of cells differ significantly within the contingency table 2×7 (2 groups of users and 7 clusters), post-hoc tests were performed. Significant differences between the number of predictions associated with efficient and inefficient user movement patterns have been identified within Cluster 2 ($\chi^2(1) = 36.97$, $p < 0.0001$), Cluster 3 ($\chi^2(1) = 19.91$, $p < 0.0001$), Cluster 4 ($\chi^2(1) = 5.55$, $p < 0.01$), and Cluster 6 ($\chi^2(1) = 25.59$, $p < 0.0001$). No specialisation has been identified for Cluster 1, 5 and 7.

These results suggest that Clusters 2 and 3 reflect good rules or good navigational strategies, while Cluster 4 and 6 represent inefficient rules. In the following sections, symbolic rules will be associated with these clusters, and their interpretation will be described in detail.

Cluster 1 does not show any specialisation, suggesting that the rule which it represents has been equally employed by both efficient and inefficient navigators. This is because it consists of a basic spatial behaviour which is worth mentioning. Cluster 1 groups the segments of trajectories performed in the very initial stage of exploring or revisiting a level. The user locations are usually within the perimeter of the virtual elevator or just in front of it. The movements consist of a set of translations from the starting point, usually in the direction of the initial heading (e.g. North), ending always with rotations involving considerable changes of heading. In other words, this cluster suggests the following rule: from an initial position, move 1–2 metres in the direction of the heading and then perform rotations, to acquire knowledge of the spatial layout and in particular about the nearest landmarks.

Cluster 2 groups the best predictions which occurred on the ground and second floor. This cluster has been identified as representing a good rule and offers a generalisation of the rule associated with Cluster 1. Similar to Cluster 1, Cluster 2 encompasses movement patterns close to the starting point, but an additional feature can be identified. The translations are performed in the major open area located closer to the starting point, area which can expend until some remarkable landmarks can be identified (e.g. almost until the middle of the bigger room). The segments of trajectories which have been grouped by this cluster belong to the ground floor or level 2, but not to the level 1 or 3, which do not offer an open area for such an exploration.

Some of the user's actions consist of translations towards the North direction which are followed by rotations. Such changes of heading allow the users to acquire a better view field. This behaviour is an extension of the one suggested by Cluster 1. The difference is that these rotations are performed on a location closer to the centre of the bigger room. Actually the centre of this larger open area acts as a *virtual* landmark whose attractiveness, based on its location, resides in the largest overview that the user is enabled to acquire.

Cluster 3 groups the best predictions which occurred on the first, second and third floor. This cluster groups segments of trajectories performed by the users with high performance in the search task, performance which could be reached only on the basis of good navigational strategies. Cluster 3 groups the best predictions related to the exploration of the middle area of the small room, located on the right hand side of the spatial layout. The levels associated with this cluster are all but the ground floor, which is populated with several landmarks in the area targeted by this cluster. This aspect impedes users to carry out a good coverage of it.

Within Cluster 3 the users' movements, usually heading towards South-West, are translations performed from the upper part of the room towards the middle of it, or from the middle towards the lower part. Interestingly, this movement patterns resemble only vaguely the wall following behaviour, typical for exploration. It seems, that the currently employed strategy allows users to cover the space more efficiently, according to the energy conservation principle, as compared to the wall following technique.

Actually, the users seem to navigate somewhere on the median space between the nearby landmarks or between the nearby landmarks and the room walls. This trajectory course keeps open several options, since the user can move towards any of these landmarks, with minimum energy consumption. Such movement patterns require only small translations but considerable changes of heading.

Cluster 4 groups the best predictions which occurred on the ground, first and third floor. This cluster has been identified as representing a poor rule. Cluster 4 groups the segments of trajectory encompassing two referential landmarks: the entrance to/exit from the virtual elevator and the door between the big room and the small room from the right hand side of the spatial layout. The actions within this cluster represent entrances from the big room to the adjacent smaller room. The area from which these movements are originated is approximately 3 metres further from the separating wall between these two rooms.

Each door in the virtual building is a sliding door, which means that it opens when the user is in its proximity, such as 1–2 metres, and closes when the user is further away. In addition, the intention to pass through the door as opposed to passing by the door, should be indicated by the heading. The closest the user's heading is to the orthogonal direction on the door, the higher the probability that the door will open. The door remains close when one's heading is parallel with the door, even when one's location is very close to the door. Moreover, going through the narrow passage of the door, within a limited time frame, involves additional cognitive overload which impedes the performance of low spatial users. This kind of door requires skilled users whose eye-hand coordination works well, who can estimate both the distance and the heading needed for the door to open, and who can therefore anticipate the moment of opening. Only through mastering these skills, one could pass the door smoothly and after the first attempt.

Within this cluster, the users' heading is usually towards South and is changing in order to adequately approach the door. However, the consecutive changes of users' heading suggest continuous readjusting of their orientation, since the door seems to be approached from a wrong angle. These changes of heading occur usually in the vicinity of the door (i.e. 1–2 metres), when in fact one should look for the proper heading a bit earlier, and maintain the constant heading without repeatedly changing it.

Another group of movement patterns belonging to this cluster involve successive changes of heading after the user has entered the smaller room. In other words, they represent attempts to acquire a larger view field immediately after the user has arrived in a new room, confirming the behaviour identified in Cluster 1. However, this behaviour has been further refined by high spatial users (see Cluster 2) who decided to choose a better location for performing such heading changing, namely in the main open area of the current room. A set of consecutive rotations performed in such a place provides a better view field, increasing the information about the room spatial layout. This enables user to see the entrance point from a different position, which help integrating it in the spatial representation.

Cluster 6 groups the best predictions which occurred on the ground, first and second floor. This cluster has been identified as representing a poor rule and is a mirror cluster of the Cluster 4. Therefore, the observations made above apply to this cluster as well.

4. DISCUSSION

Much research has been focused on investigating those characteristics of VEs which impact on navigation training and in particular on the effectiveness of transfer from VEs to the real world (for a review, see Waller et al. ([12])). This line of research is primarily concerned with manipulation of the technical aspects of VE design such as fidelity or realism of the interface, media, quality of VE and training time, or presence of additional navigational cues [10, 1, 2, 13, 11].

Such tools, among which the map is the most common, are easy to place on a VE and they seem to impact significantly on the quality of the acquired spatial knowledge [1]. However this approach to training navigation has a major limitations. In all these studies, the navigational tools are applied equally to each end user, without any attempts to tailor their design to user profiles. Such user profiles could be grounded on user mental models of navigation.

The difficulties of investigating spatial mental models and the limitations of techniques developed for this purpose explain the lack of studies in this area. This is the gap that this work tries to address. Its major contribution is the proposal of machine learning techniques to overcome the limitations of traditional methods for eliciting such models. It focuses on investigating user mental models of navigation, in order to build a user model of navigation The user model consists of a set of navigational rules which can support efficient spatial behaviour, and accordingly can be exploited for designing VEs for navigation assistance. The proposed methodology allows the identification of two efficient procedural rules:

- *Surveying zone* From the starting point, the users search for the nearby largest open area from where a larger view field enables them to acquire information about both spatial layout and landmark configuration.

- *Median paths* As long as none of the landmarks acts as a strong attractor, in other words none of the landmarks raises the user's interest, he/she will follow an *equilibrium path* between the nearby landmarks (this could also include walls). This rule resembles the "wall following" technique but it is more efficient. As soon as one of the landmarks captures user's interest, the equilibrium path is not followed anymore and the user gravitates towards this particular landmark, with minimum energy expenditure.

The inefficient procedural rule is related to the difficulties encountered by low spatial users in passing through the sliding doors. Through their design, these doors put unusual demands on these users. This rule is not primarily related to navigation, but suggests once again the importance of individual differences in designing VEs. These users lack some skills and in order to help them overcoming their limitations, the doors should be designed differently.

The major challenge of this work was the data mining process based exclusively on user behavioural data. This data should not only be understood but also interpreted in terms of navigational rules and strategies. Another challenge consisted of tuning the developed methodology in such a way to increase its potential for capturing navigational rules which discriminate best efficient from inefficient spatial behaviours. However, the demand of using this methodology in real-time requires it to run with relatively limited computational resources.

The benefits of this work can be seen at both theoretical and practical level. At a theoretical level, the investigation of spatial mental models enriches the understanding of how humans perceive the space, make sense of space and exploit it. Apart from the theoretical contributions which such an understanding enables, the practical ones could lead to increased usability of VEs, through identifying ways to support inefficient spatial behaviour and challenge the efficient one. This methodology is objective, involving the analysis of user behaviour rather than his/her subjective thoughts. It is also unobtrusive, since it requires no additional involvement of the users, and even more importantly, it can be employed during task completion. The dynamics of user's behaviour can be captured online, a fact which ensures system's capacity to dynamically adapt to the user's navigational model.

Increasing the percent of best RNN predictions included in the rule extraction module could lead to a larger set of navigational rules. Additionally, increasing the sample size of study participants could provide more reliable results. An interesting study direction to be followed consists of extracting such rules, from trajectory paths performed in completely different VEs, where different variables regarding spatial layout and landmark configuration can be efficiently controlled and manipulated. Such study will help refining the currently extracted rules.

5. REFERENCES

[1] R. Darken and J. Sibert. Navigating large virtual spaces. *International Journal of Human–Computer Interaction*, 8(1):49–72, 1996.

[2] R.P. Darken and W.P. Banker. Navigating in natural environments: A virtual environment training transfer study. In *Proceedings of VRAIS '98*, pages 12–19, 1998.

[3] R. Ellis and G.W. Humphreys. *Connectionist Psychology: A Text with Readings*. Psychology Press, Hove, 1999.

[4] J.L. Elman. Finding structure in time. *Cognitive Science*, 14:179–211, 1990.

[5] T. Ghiselli-Crippa. *Spatial and Temporal Factors in the Acquisition of Spatial Information: A Connectionist Model*. PhD thesis, School of Information Sciences, University of Pittsburgh, 2000.

[6] S. Kaski. *Data Exploration Using Self-Organizing Maps*. PhD thesis, Department of Computer Science and Engineering, Helsinki University of Technology, 1997.

[7] T. Kohonen, J. Hynninen, J. Kangas, and J. Laaksonen. SOM PAK: The self-organizing map program package. Technical Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science, 1996.

[8] K. Plunkett and J.L. Elman. *A Handbook for Connectionist Simulations*. MIT, Cambridge, MA, 1997.

[9] A. Psarrou and H. Buxton. Motion analysis using recurrent neural networks. In *International Conference in Artificial Neural Networks*, 1994.

[10] G. Satalich. Navigation and wayfinding in VR: Finding the proper tools and cues to enhance navigational awareness. Master's thesis, Department of Computer Science, University of Washington, 1995.

[11] D. Waller. Individual differences in spatial learning from computer—simulated environments. *Journal of Experimental Psychology: Applied*, 8:307–321, 2000.

[12] D. Waller, E. Hunt, and D. Knapp. The transfer of spatial knowledge in virtual environment training. *Presence: Teleoperators and Virtual Environments*, 7(2):129–143, 1998.

[13] D. Waller and J. Miller. A desktop virtual environment trainer provides superior retention of a spatial assembly skill. In *Proceeding of ACM SIGCHI 98 (Poster)*, 1998.

[14] A. Zell, G. Mamier, M. Vogt, N. Mache, R. Hubner, S. Doring, K.U. Herrmann, T. Soyez, T. Schmalzl, T. Sommer, A. Hatzigeorgiou, D. Posselt, T. Schreiner, B. Kett, G. Clemente, and J. Wieland. *The SNNS Users Manual Version 4.1*. Institute for Parallel and Distributed High Performance Systems, 1995. URL: `http://www-ra.informatik.uni-tuebingen.de/SNNS/` (version current as of 4th July 2002).