Adaptive Motion Pattern Recognition: Implementing Playful Learning through Embodied Interaction

Anja Hashagen, Christian Zabel, Heidi Schelhowe, and Saeed Zare

University of Bremen, TZI, dimeb (Digital Media in Education),
Bibliothekstr. 1, 28359 Bremen, Germany
{hashagen,chr,schelhow,zare}@tzi.de

Abstract. The concept of embodiment plays an emergent role in Human-Computer-Interaction. Accordingly, we conceptualized, implemented, and evaluated an adaptive motion pattern recognition system for an educational installation called *Der Schwarm*. We implemented three algorithms and compared correctness and processing speed. *Der Schwarm* aims to encourage children to learn about technology and interprets free body movements. The motion pattern recognition system fosters embodied playful learning, as an evaluation with children shows.

Keywords: Motion Pattern Recognition, Playful Learning, Embodied Interaction, Children Education, HCI, Virtual Environments.

1 Introduction

Kevin is 11 years old and on a study trip with his class. The teacher only announced a visit at local University. Now the students take place around a marked rectangle on the floor, in which a swarm of strange-looking little bugs move. After a short welcome by two researchers, Kevin slowly enters the *area and the bugs immediately approach him. He is surprised and runs to a corner of the rectangle, while the swarm follows him. "Cool, they can see me!" he shouts. Kevin quickly waves his arms and the swarm backs off and attempts to escape. "Step on them! Catch them!", "Look, they are not real, but made of light", and "Run away and see what they do!" add some classmates. The class is very exited and curious about the artificial swarm, where it comes from, why it can see Kevin, and how he may control the swarm. After Kevin left the rectangle and some of his classmates interacted with the swarm, the class discusses upcoming theories about functionality and underlying concepts and creates test scenarios to proof their hypotheses.*

This scenario describes a typical workshop with children and the installation *Der Schwarm* [6] (see System Implementation *Der Schwarm*) conducted by our research group). From our experience in the field of Digital Media and Education free body interaction provides a good starting point to pique children's curiosity about technology. Therefore, *Der Schwarm* implements the concept of Embodiment after

V.G. Duffy (Ed.): Digital Human Modeling, HCII 2009, LNCS 5620, pp. 105–114, 2009.

Dourish [4] and offers the possibility to interact with a computer on a bodily level. The abstract functionality regarding technical setup and swarm algorithms can be explored playfully through the medium body. The scenario emphasizes children's motivation that is considered as a key factor to gain a deeper understanding, following the principle of Playful Learning [9]. In order to support the process of exploring abstract technological concepts through concrete body interaction driven by motivation, we implemented motion pattern recognition and integrated the software to the installation *Der Schwarm*.

2 Motion Pattern Recognition System

Related work mainly exists in the field of gesture recognition. Rubine developed a gesture recognition system, which employs Fisher's Classification Functions to recognize patterns with devices such as mouse and stylus pen [11]. A commercial solution is the software iisu™ by Softkinetic. A 3D depth sensing camera tracks body movements with arms, legs etc and the gesture is recognized by a comparison of tracked position information with sample gestures [8]. Another example of a gesture recognition system is the game console Wii by Nintendo with the input device Wiimote [13].

The motion pattern recognition system we developed differs from these solutions in its input and tracking method. Our objective is to recognize motion patterns, more precisely the recognition of a walked path on the floor instead of gesture patterns, drawn paths in space. Anyway, the classification of the patterns requires similar conditions and can be achieved with similar methods.

Within our system, the user enters pre-trained motion pattern by walking in a monitored area with the ratio of 4:3. The user's position is tracked by a laser scanner. A visual reaction is triggered after a motion pattern is detected and projected on the floor within the monitored area. The patterns are grouped to sets of similar motion patterns such as *Geometry* and *Numbers*. The system is adaptive and therefore provides a training function, which lets the user enter new motion patterns by walking patterns repetitively.

Two of the three classification algorithms we implemented are based on Discriminant Analysis. This affects the general procedure of training and classification and therefore, we introduce the concept before describing training or classification details.

Generally, after having trained a new motion pattern, the pattern is represented as Discriminant Function [3]. For classification the same function is calculated to distinguish motion patterns and results in a discriminant value (Equation 1). The higher the resulting discriminant value, the higher is the similarity of the patterns and the motion pattern is classified.

$$D = k_0 + k_1 * m_1 + k_2 * m_2 + \ldots + k_n * m_n . \quad (1)$$

Equation 1 shows Fisher's Discriminant Function, a linear combination that calculates the distinction of two patterns. Variables m_1 - m_n represent characteristics of the motion pattern entered. During classification the characteristics are inserted to

the function. The coefficients k_0 - k_n, calculated with matrix manipulation, are based on the results of the training. With the characteristics (from data input) and the coefficients (inserted during training) the function is calculated and D represents the resulting discriminant value.

Classifications of more than two patterns require several discriminant functions. The number of required discriminant functions depends on the number of motion patterns and on the discriminant analysis variant. In [7] and [12] are two variants mentioned, Fisher's Classification Functions and Fisher's Linear Discriminant Function. We employed both as algorithms within the motion pattern recognition system.

2.1 Motion Pattern Training

Training is the process to create new motion patterns. Several input repetitions are necessary to ensure an adequate recognition rate. Besides high recognition rates our aim is to reduce computation complexity and have a minimum of needed repetitions. From our experience about 15 repetitions lead to reasonable classification results. We discuss the training complexity and recognition rates in the technical evaluation chapter.

During training the average value of each characteristic is calculated to adjust the coefficients of Discriminant Functions. The adjustment ensures a stable motion pattern classification with Discriminant Analysis and the third algorithm developed by the authors.

2.2 Motion Pattern Recognition

After motion patterns are trained and known to the system, software needs to recognize such patterns based on the walked and tracked path of the user during his/her data input. The recognition of motion patterns requires such as gesture pattern recognition the computation of several components. Therefore, we differentiated the pattern recognition process to the following sequential steps.

1. Motion point measurement (data input)
2. Characteristics extraction
3. Classification
4. Reaction triggering

During motion point measurement the path walked by the user is tracked and saved. Afterwards the pattern characteristics are extracted and compared with pre-trained motion patterns during classification. If the pattern is recognized an appropriate reaction is triggered, otherwise a text message is shown.

Motion Point Measurement. The user’s position is tracked with a laser scanner and a sequence of two-dimensional points is extracted to recognize a motion pattern. During the input of the pattern the user’s coordinates are constantly measured and saved. The recognition of motion patterns we developed is based on the calculation of motion points. The average of five successive coordinates results in one motion point. Sequences of these motion points define a motion pattern. Figure 1 shows the

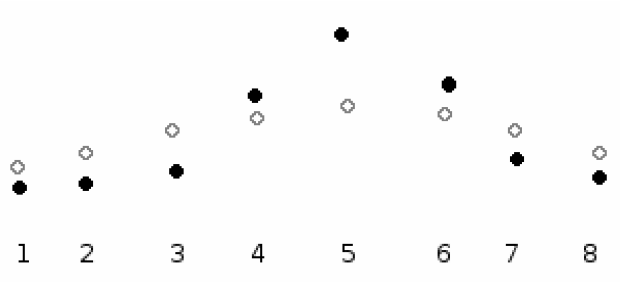

Fig. 1. Coordinate sequence (black) and Motion Point Sequence (white)

comparison of two sequences, the original coordinate sequence (black) from the laser scanner and the motion point sequence (white) after the average value calculation. The average value of e.g. motion point no 3 is calculated from coordinates no 1-5.

Calculating the averages leads to a better performance during the classification and smooth edges. Hence, irregularities and small errors in the motion pattern are reduced to ensure a stable recognition.

Characteristics Extraction. Based on the motion point sequence, a characteristics vector is calculated which specifies essential properties for the recognition. The vector describes the characteristics of the pattern.

The selection and the number of the characteristics influence quality as well as performance of the recognition process. In order to ensure an adequate performance the complexity in calculation and thus the number of characteristics has to be as low as possible. However, a high recognition rate requires as much characteristics as possible. We considered quality, performance, and hardware configuration and an absolute minimum of four characteristics is required. In our experiences, about ten characteristics guarantee a stable recognition (see Evaluation).

The chosen characteristics ensure an invariance of scaling, thus, the size of entered motion patterns is irrelevant. Furthermore, we only consider position relations instead of absolute positions, which lead to translation invariance.

Classification. After the motion points are extracted and at least eight motion points were found the classification process starts. Based on the characteristics vector, the motion pattern can be classified. A successful classification result is a motion pattern class, whose characteristics are most similar to the characteristics of the entered motion pattern.

We implemented three algorithms and compared recognition rate and classification complexity. Two algorithms employ the above mentioned Discriminant Analysis, Fisher's Classification Functions (FCF) and Fisher's Linear Discriminant Function (FLDF) [7] [12]. The third algorithm implements Average Classification Function (ACF), which is developed by the authors of this paper and calculates the average values of the feature vector's elements.

Fisher's Classification Functions (FCF). This variant of Discriminant Analysis was firstly described by Fisher [12] and also calculates one discriminant function for each class (Equation 1). Finally, the class with the largest discriminant value is chosen [7].

This version of Discriminant Analysis was successfully applied in [11] and achieved a gesture recognition rate of 96%. The differences to our classification algorithm are marginal and just differ in the input method, since the further processing is similar.

Fisher's Linear Discriminant Function (FLDF). Fisher's second algorithm, which is associated with motion pattern recognition, generates one discriminant function (Equation 1) to compare two motion patterns. Following the principles of FCF, FLDF calculates a discriminant value and in doing so determines the motion pattern with the highest similarity [12].

Furthermore, Elpelt described a method to distinguish several motion patterns with FLDF, which requires one function for each pair of motion patterns (entered and system-known) [5]. A motion pattern is recognized if each discriminant function returns a positive result.

Average Classification Function (ACF). In order to gain a better performance we developed a function, which does not calculate matrices during training and classification unlike FCF and FLDF (Equation 2). The algorithm calculates average values and standard deviations of each characteristic.

$$C = C + 1, \text{ if } s_x \geq |a_x - m_x| \text{ , for all x in n .} \tag{2}$$

In Equation 2 the result C is the classification value of the actual system-known motion pattern, s_x represents the standard deviation of the actual characteristic, and a_x the average value. M_x is the characteristic of the entered and hence to be classified motion pattern. If the difference of the actual characteristic to the system-known motion pattern's average value is equal or lower than the standard deviation, the classification value is incremented. This calculation is performed for each characteristic and results in one classification value per system-known motion pattern. The result of the classification and therefore the pattern recognition is the motion pattern with the largest value, which represents the highest similarity.

Reaction Triggering. If the motion pattern was classified successfully, a predefined reaction is triggered. We implemented a set of instant visible reactions as examples. These reactions are strongly related to the field of application and the installation Der Schwarm. Depending on purpose, specific usage and integration to other software systems, all sorts of reactions can be implemented to our adaptive motion pattern recognition system.

3 System Implementation at Der Schwarm

The adaptive motion pattern recognition has been integrated to the multi-agent-system *Der Schwarm* (translation: the swarm, the flock), which allows free body movement interaction with a virtual swarm. The first version of the installation was planned and implemented in 2004[1], continuous enhancements in technology and workshop

[1] First idea, concept, and realization particularly by Merten Schüler and Andreas Wiegand.

concept have led to the actual installation. Anyway, the employed algorithms computing the motion pattern recognition can easily be used in other scenarios.

Der Schwarm is a technological learning environment consisting of hardware and software that detects, tracks, and interprets free body movements. The hardware of *Der Schwarm* consists of a laptop, a laser scanner, and a projector. The laser scanner is installed at table height and detects free body movements and returns a continuous stream of two-dimensional position information about the interacting person to the laptop. The software computes a reaction to this technical representation of body movements and finally the projector produces a visual feedback. The projector is installed above the interacting person and describes a projection area of at least 6.0m x 4.5m (ratio 4:3) depending on its installation height. The laser scanner is calibrated to track movements within the projection area. Figure 1 shows the hardware setup of *Der Schwarm.*

The system's reaction to free body movements is computed by special software and is visualized as a flock of light spots. Reynolds' solution for steering autonomous characters is employed to simulate swarm behavior [10]. Besides the implemented swarm behavior, the light spot's steering direction, velocity, as well as appearance is influenced directly by the interacting person's movements. We developed six states and created parameter sets for the swarm representing the behavior patterns trust, curiosity, observance, escape, confusion, and aggression. The changing parameters are for instance level of herd instinct, basic velocity, and basic distance to interacting person. Both behavior and color of the flock changes with its state, so that a flock of light blue, slowly moving, and the interacting person closely following light spots represents curiosity, whereas a red, quickly moving, and the interacting person chasing flock represents aggression.

Since we developed a flexible software structure, new image sets of light spots are easily interchangeable and shapes (and colors) such as bugs, fish, dragonflies, circles, and squares provide room for experimentation. The image set shown in Figure 2 has been used in several workshops with children conducted by our research group.

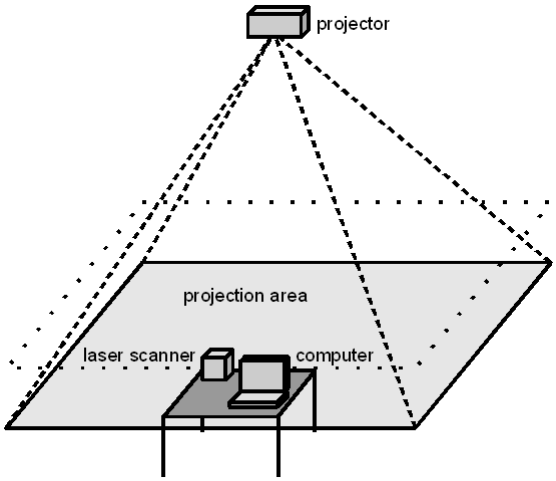

Fig. 2. Hardware setup of Der Schwarm [6]

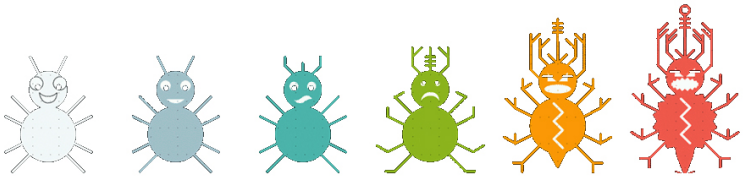

Fig. 3. Image sets of light spots for every state

Altogether, the installation *Der Schwarm* provides a technical starting point to foster children's motivation to learn about technology. The system's reaction to free body movements provides room for interpretation and discussion, as our experience from several workshops with children at the age of 9-14 years shows. The interacting person and the participants interpret technical output information semantically and curiosity about underlying techniques arouses. A didactical concept is needed to provoke children's curiosity and motivation, which are requirements for learning. We developed and tested a concept [6] based on the statement of Ackermann in which she argues that a combination of interaction phases with immersion, so-called *diving in* on the one hand and reflection, known as *stepping out*, is needed in order to gain deeper understanding [1] [2].

The integration of the adaptive motion pattern recognition system to *Der Schwarm* aims at introducing abstract models with a bodily approach. Even to train new motion patterns requires advanced abstract thinking for instance to mentally compare similarity with other trained patterns, several repetitions of the path as well as decision making about the exact appearance and path to be walked.

The children immediately experience the consequences of their decisions and gain deeper understanding. In the end, the system might mistake an imprecise trained square for a circle. The finding can be technology's incapability to interpret or the need for more different looking patterns and afterwards a resulting action might be an exact training of a square pattern or the development of a new pattern. Additionally, features like translation and scaling invariance offer more possibilities in pattern classification and therefore implement extra level of abstraction.

4 Evaluation

Different levels of our work, namely system specification and Human Computer Interaction require two evaluations, a technical as well as a user evaluation. The technical evaluation aims at producing technical results about the motion pattern recognition software system, specifically about recognition rates and training complexity. For the user evaluation probands were asked to enter motion patterns to the system and comment there experiences, while we observed their actions.

Both evaluations were performed with two datasets of patterns, which we trained in advance. The number dataset contains digits from 0 to 9, while the geometry set includes geometrical motion patterns such as square, triangle, and circle. The datasets differ in pattern complexity and similarity of patterns entered within a set.

4.1 Technical Evaluation

Both system-known datasets were tested with each of the three implemented algorithms. We entered every pattern of each dataset and calculated the average recognition rate per algorithm. Incorrect classification or no recognition is interpreted as an error. Additionally, we calculated the theoretical complexity in calculation during training and classification. Table 1 shows the recognition rates and complexity notations of the above mentioned algorithms. C represents the number of classes and F the constant number of features for each method.

Table 1. Recognition Rates of Three Classification Functions

Algorithm	Number Dataset	Form Dataset	Average	Complexity	
				Training	Classification
FCF	94%	89%	91,5%	$O(F^2 + C)$	$O(C)$
FLDF	50%	100%	75,0%	$O(F^2 + C^2)$	$O(C^2)$
ACF	66%	77%	71,5%	$O(F + C)$	$O(C)$

In average, the implementation of FCF achieves the best recognition rates with a very good consistency. The FLDF algorithm achieves better results than ACF, but the recognition rates are inconsistent and some motion patterns could never be classified correctly, whereas other patterns were always recognized. The recognition rates with ACF are more consistent than FLDF and show about the same average recognition rate as FLDF.

FLDF shows the highest theoretical calculation complexity in training as well as classification. FDF and ACF have a similar classification complexity. FDF shows a mid-range complexity in training, but as we stated, has high recognition rates. The best complexity rates in training shows the ACF algorithm and has potential for improvement to increase recognition rates at a good performance.

4.2 User Evaluation

The second part of the evaluation aims at gaining information about the software's usability and accessibility for user respectively actors with the system. Therefore nine probands were asked to enter pre-trained motion patterns. Since the main target group of the installation *Der Schwarm* are children, we focus on four probands who were aged between 9-15 years.

Every proband was asked to enter each motion pattern of the number and geometry dataset. We prepared printouts with images of the motion patterns to prevent misunderstandings regarding the notation of for instance digits such as 7, 4, and 1. During the interaction with the system, we observed the proband’s actions and interviewed them afterwards. Focus of interest was the level of difficulties entering the motion pattern and the motivation to succeed. Since we consider motivation as a key factor in children’s process of understanding, these observations are crucial for further developments of the motion pattern system and its integration to *Der Schwarm*.

The results of the evaluation with probands were predominantly positive. In general, the probands managed the task well and showed a high interest in the system as well as the underlying concepts. The majority of the motion patterns were entered without problems and successfully classified (based on the recognition rates of the FCF-algorithm). Some probands struggled with their orientation during walking in the monitored area, thus a repetition was necessary to enter a motion pattern successfully. The lack of orientation mainly occurred, if a complex pattern required a crossing of the walked path, for instance at the digits 9 and 4.

Altogether, especially the important group of children and youth showed a high motivation and tried to enter the different motion patterns several times. Software enhancements, for instance a visualization of the actor's walked path will probably help to solve problems observed.

5 Conclusions and Future Work

We developed an adaptive motion pattern recognition system that detects two-dimensional motion patterns walked by a person in a laser scanner monitored area. New patterns can be trained with few repetitions. The best classification rate of motion patterns is achieved by Fisher Classification Functions, as the comparison of three algorithms shows. High potential for improvement has the Average Classification Function, we developed, since the algorithm shows the best results in calculation complexity.

We integrated the pattern recognition system to the learning environment *Der Schwarm*, a project of our research group. As the introducing scenario states, the motion pattern recognition system aims at supporting the general objective of *Der Schwarm* to pique children's curiosity about technology. The installation provides an environment to explore Digital Media playfully and to gain deeper understanding of underlying concepts by embodied interaction. Operating the motion pattern recognition system with *Der Schwarm* requires an intense examination of technology and its principles and can support the process from immersive engagement to abstract understanding. During the user evaluation children of the main target group at the age of 9-15 years interacted with *Der Schwarm* and the motion pattern recognition system. They showed a high motivation to learn about the technology.

Besides technical improvements in performance and classification rate, our next approaches are enhancements in usability, such as a visualization of the walked path and a rotation invariance. Furthermore, we plan the implementation of appropriate reactions to a (un)successful recognition, that supports the didactical concept in enabling children to playfully explore abstract concepts of through concrete interaction.

References

1. Ackermann, E.K.: Perspective-Taking and Object Construction: Two Keys to Learning. In: Constructionism in Practice: Designing, Thinking, and Learning in a Digital World, pp. 25–35. Lawrence Erlbaum, Mahwah (1996)

2. Ackermann, E.K.: Constructing Knowledge and Transforming the World. In: Tokoro, M., Steels, L. (eds.) A Learning Zone of One's Own: Sharing Representations and Flowing Collaborative Learning Environments, pp. 15–37. IOS Press, Amsterdam (2004)
3. Backhaus, K., Erichson, D., Plinke, W., Weiber, R.: Multivariate Analysemethoden - Eine anwendungsorientierte Einfuehrung. Springer, Berlin (2003)
4. Dourish, P.: Where the Action Is: The Foundations of Embodied Interaction. MIT Press, Cambridge (2001)
5. Elpelt, B., Hartung, H.: Multivariate Statistik. Oldenbourger Wissenschaftsverlag, Munich (2007)
6. Hashagen, A., Schelhowe, H.: 'Der Schwarm" - Playful Learning with an Interactive Virtual Flock. In: Auer, M.E. (ed.) Interactive Computer aided Learning (ICL) International Conference Villach/Austria 2008, The Future of Learning – Globalizing in Education. Kassel University Press, Kassel (2008)
7. Klecka, W.A.: Discriminant Analysis. Sage Publications, Thousand Oaks (1980)
8. Softkinetic, S.A.: Product Datasheet IISUTM —iisu.pdf (2008), http://www.softkinetic.net/Files/media/Softkinetic-Product-Datasheet-iisu.pdf
9. Resnick, M.: Edutainment? No Thanks. I Prefer Playful Learning. In: Associazione Civita Report on Edutainment. pp. 1–4 (1987)
10. Reynolds, C.W.: Steering Behaviors for Autonomous Characters. In: Proceedings of the Computer Game Developers Conference 1999, pp. 763–782. Miller Freeman Game Group, San Francisco (1999)
11. Rubine, D.: The Automatic Recognition of Gestures. PhD thesis, Carnegie Mellon University (1991)
12. Srivastava, M.S.: Methods of Multivariate Statistics. John Wiley & Sons, Inc., New York (2002)
13. WiiTM, Nintendo®, http://wii.com/