A Tennis Training Application Using 3D Gesture Recognition

Cristian García Bauza[1,2], Juan D'Amato[1,2], Andrés Gariglio[1], María José Abásolo[3,4], Marcelo Vénere[1,5], Cristina Manresa-Yee[6], and Ramon Mas-Sansó[6]

[1] Instituto PLADEMA, Universidad Nacional del Centro de la Provincia de Buenos Aires, Tandil, 7000 Argentina
[2] CONICET, Argentina
[3] Universidad Nacional de La Plata, La Plata, 1900 Argentina
[4] Comision de Investigaciones Científicas de la Provincia de Buenos Aires, Argentina
[5] CNEA, Argentina
[6] Universitat de les Illes Balears, Palma, 07122 España
{crgarcia,jpdamato}@exa.unicen.edu.ar, andres.gariglio@gmail.com, mjabasolo@lidi.info.unlp.edu.ar, venerem@exa.unicen.edu.ar, {cristina.manresa,ramon.mas}@uib.es

Abstract. This paper presents a sport training system which recognizes user movements from data of the Wiimote device with accelerometer technology. Recognizing a new gesture involves the normalization of the Wiimote data and searching in a gesture templates database. The Dynamic Time Warping (DTW) comparison algorithm is used as a correlation function to compare the new gesture with every template. Based on prior training, the system can successfully recognize different sport shots. Particularly the system is instantiated for tennis training. The user visualizes the trajectory of the ball in a three-dimensional environment and he can interact with virtual objects that follow Newton dynamics.

Keywords: Gesture recognition, Wiimote, Videogames, Sport training, Tennis training, Accelerometer, Computer graphics, Dynamic Time Warping.

1 Introduction

The keyboard, the mouse and the monitor continue being the standard configuration for the human-computer interaction (HCI). Nevertheless, during the last years diverse technologies have appeared that try to change the way how users interact with the computer and other electronic devices. Recently there have been big advances that allow the use of 3D gestures to interact in virtual environments, doing tasks as selection, manipulation, navigation and system control [1]. Though great quantity of developments exists focused on recognizing user gestures, most of them use conventional tracking devices [2][3][4][5]. The recognition of gestures from information given by gyroscopes or accelerometers is an emerging technology in HCI [6]. With the rapid development and mass-production of the technology in mobile

F.J. Perales, R.B. Fisher, and T.B. Moeslund (Eds.): AMDO 2012, LNCS 7378, pp. 239–249, 2012.

phones and video consoles, people have access to one or more devices equipped with an accelerometer, for example, the Apple iPhone or the Nintendo Wiimote. These devices with wireless capability offer new possibilities to interact and to be used in a wide range of applications, such as domestic appliances control, home automation, special education, rehabilitation, augmented reality and sports training.

The present work focuses on a sport training system that recognizes 3D gestures by analyzing the Wiimote's accelerometer information. The application uses 3D technology to represent game scenes and simulate the physical behavior of virtual objects that can interact with the ball. User movements are compared to previously generated templates. The developed tools can help sports trainers to detect the weaknesses and strengths of each user. Particularly the system is instantiated in a tennis training application that is called *WiiSimTennis.*

The paper is organized as follows: section 2 reviews the literature on sport training systems; section 3 describes the gestures recognition process of our system; section 4 describes the system architecture and interaction; section 5 presents some experiments and results with the tennis training application and finally section 6 presents the conclusions and future work.

2 Related Work

Successful performances in sports games require not only efficient and correctly executed techniques but also a high level of perceptual skills. Müller et al [7] comment that sports training aims at minimizing the mistakes to improve the performance-precision relationship. This demands a complete analysis of the athlete's movements, who will then receive a feedback on how to improve the technique or information to compare his or her performance with previous evaluations.

To be able to evaluate specific stimuli is necessary to be able to control the training conditions. Commonly video presentations or presentations on natural size screens are used. In [8] the behavior of the handball archers was analyzed. In [9] video is used to carry out visual searches of a football player's strategies. Clearly, the training systems based on three-dimensional virtual environments are better than the video-based, since they give more versatility in the creation of scenes and situations for training [10][11][12][13].

Trained user immersion is enhanced by using a device such as the Wiimote and by virtually recreating the user gestures on a screen [14]. Siemon et al [15] show that sports like bowling (or tennis) can take advantage of this kind of virtual reality systems to obtain a basic trainer. Although the player performance is not analyzed in a professional way, the user receives information across the graphs on the screen (for example, a curve of his forehand shot or the speed achieved in his service). Principally the beginners can obtain improvements in their performance by receiving feedback in every shot and trying it repeatedly.

3 Gesture Recognition

The Wiimote is a part of the Nintendo Wii video games console, but it can be used as a PC input device by means of a Bluetooth connection. This device is a wireless control with form of remote control which is hold by the user's hand. It allows the detection of movement in the 3D space. It includes an accelerometer that gives accelerations in three ordinal axes every certain interval of time. It also includes an infrared built-in camera that can provide information about its position, including the distance up to the PC. However, to use the position information, a more complex configuration is needed not being viable for the recognition of gestures in general cases. For this reason our work is only based on the accelerometer information.

The typical movements in sports, for example a drive or reverse shot in tennis, are what we can call a gesture. In this work we represent a shot or gesture as a movement in a certain direction in the three-dimensional space with a certain speed.

In this section we explain the fundamentals of the gestures recognition software module of our system, where the gesture concept means a sport shot or a movement.

3.1 Gesture Description

The information provided by the Wiimote accelerometer must be compiled and interpreted. The gestures realized with the Wiimote are described with three acceleration values (a_x, a_y, a_z) at every instant of time, each value corresponding to acceleration in *X, Y* and *Z* axis respectively.

There exists more than one way of representing it. One way could be a 3D path that recreates the movement performed by the user in the 3D space. To obtain a realistic curve, additional information such as position and initial orientation are necessary, but the results are not always the desired ones. Several works such as in [16] propose to carry out an estimation of the position in the 3D by means of Kalman's filters. This is effective to follow the path of the Wiimote, but fails to detect the user's movement.

In this work, we interpret the Wiimote data as three acceleration curves *Ax*, *Ay* and *Az*, corresponding to accelerations in *X, Y* and *Z* axis respectively at each instant.

It is easy to see that the curves of a gesture can be performed with different amplitude and different time or speed by different users. Leong et al. [17] observed that for a particular gesture (e.g. to draw a circle) different speeds of execution were obtained for different users and small variations were found in samples done by the same person. Due to this, the number of obtained values increases if the gesture is performed slower, whereas the number decreases if the gesture is faster than the average speed. These facts can obstruct the compilation of information and the later comparison of two gestures.

To solve the differences in speed, the sequence of information is normalized in a normalized interval of time equal to the total duration of the gesture. We follow the one proposed by Leong et al. [17] that use 50 values for the representation of every gesture. The fixed-size vector of (a_x, a_y, a_z) data is obtained by linear interpolation, and is what we call a sample. Fig. 1 shows an example of the two sets of X acceleration data: the original one with 40 values and the re-sampled or normalized one with 50 values. It can be seen that re-sampling preserves the curve of information of the original movement.

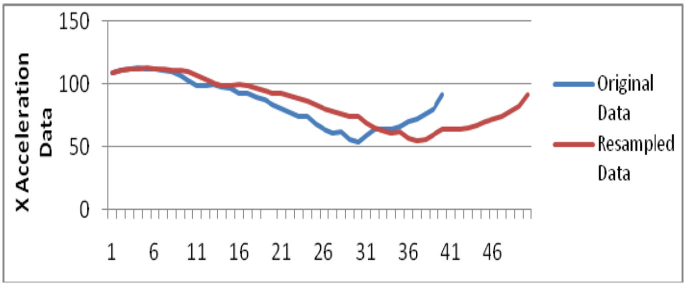

Fig. 1. Example with the original acceleration in x-axis (from time 0 to 40) and the normalized information (from time 0 to 50)

To obtain a more precise representation for every gesture we use the idea of templates such as Leong et al proposed in [17]. A template is built from a set of samples corresponding to a certain type of movement. The process of creation of a template follows the following steps. First several samples of the movement recreated by the user must be obtained; then these samples are normalized by re-sampling; finally the *Ax, Ay* and *Az* curves of the template are built with the mean value of the normalized samples at each instant of time. Fig. 2 shows an example of template built from five samples.

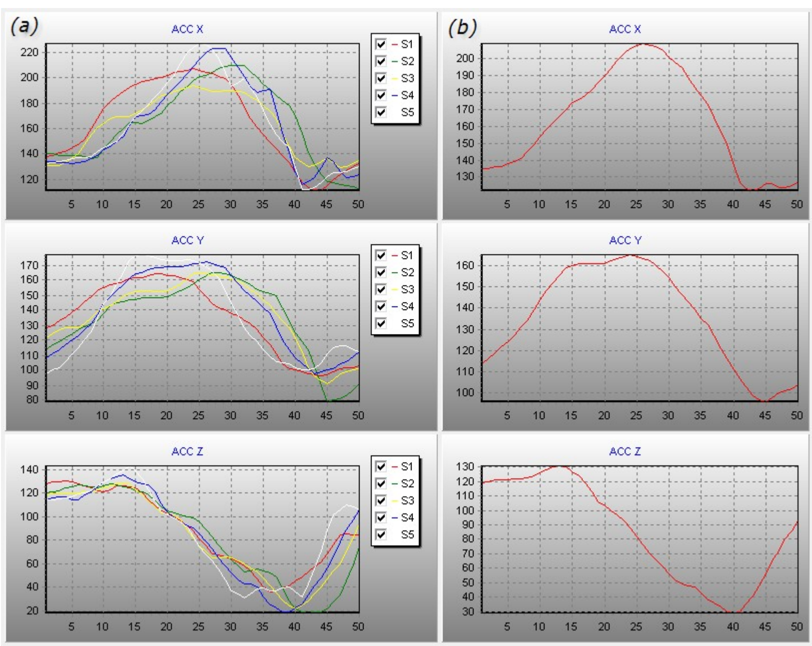

Fig. 2. Example of a template built from five samples: (a) Graphic of X, Y and Z accelerations of the five samples; (b) Graphic of X, Y and Z accelerations of the template

This template generation process can be seen as a stage of training of the system. In our system a database of the different tennis shots are built at prior. In Fig. 3 the process of creation of templates from N samples is shown. To recognize a new gesture, the vector or gesture sample is compared against the stored templates by means of the *Dynamic Time Warping* algorithm which is presented in the next section.

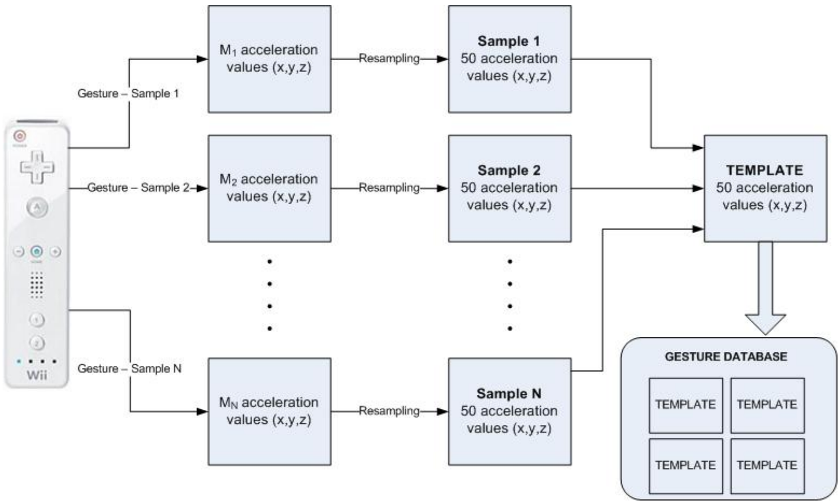

Fig. 3. Template generation process

3.2 Gestures Comparison

The *Dynamic Time Warping (DTW)* is an algorithm to measure similarity between two sequences that can change in time or speed [18]. This algorithm has gained popularity for being extremely efficient, as can be seen in [19][20][21]. It minimizes the effects of change and distortion in time, allowing elastic transformations of the series in order to detect similar forms with different phases. With an adjustment it can also be used in multidimensional gestures [22].

The utilization of the *DTW* algorithm is a correlation strategy to find the most similar template to an unknown gesture. The *DTW* determines a measure of similarity between them. The following three *DTW* values are computed between the new sample *N* obtained from the Wiimote data and every template T_i stored in the database: $DTW(N^{Ax}, T_i^{Ax})$, $DTW(N^{Ay}, T_i^{Ay})$ and $DTW(N^{Az}, T_i^{Az})$.

Senin [18] proposes to analyze the *DTW* values of every axis and if every value does not overcome a certain threshold of similarity, one concludes that the two gestures are similar.

If a new sample is compared with several templates it can happen that more than one template fulfill this previous condition. In this case the most similar template is the nearest one based on the Euclidean distance between every pair of tuplas (N^{ax}, N^{ay}, N^{az}) and (T_i^{ax}, T_i^{ay}, T_i^{az}).

4 System Description

4.1 System Architecture

The developed system is formed by different components for every task such as obtaining data from the Wiimote, recognizing sport shots, visualizing of the three-dimensional scene of the virtual environment and simulating the physical behavior of virtual objects. The graphic showed in the Fig. 4 outlines the above mentioned components and the interactions among them.

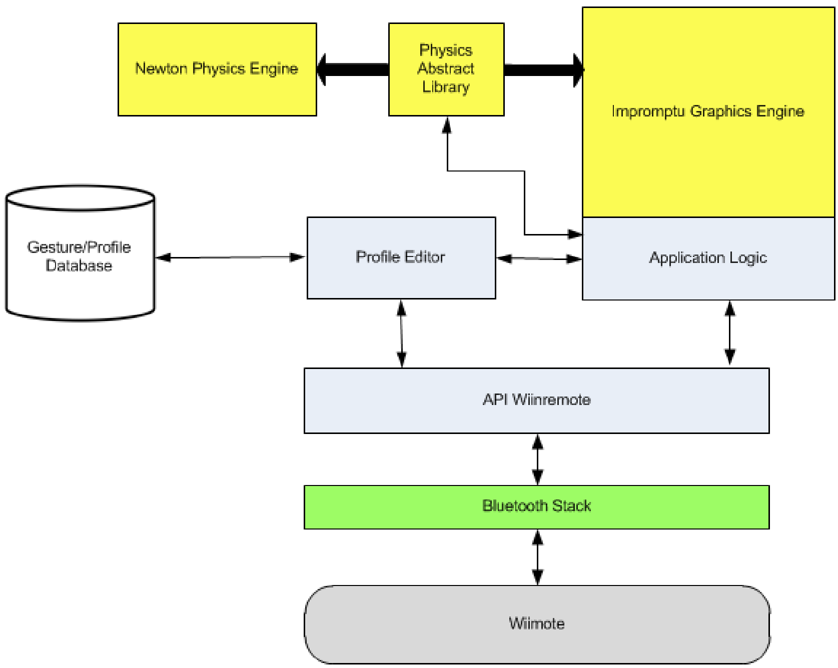

Fig. 4. System architecture

The system is based on an event-guided architecture over our Impromptu Graphics Engine [23]. This graphical engine provides a complete support to easily create and visualize three-dimensional scenes. For every scene a set of 3D models was prior created and a graphic resource administrator loads them at runtime. The profile edition module administrates the user profile, allowing the creation and edition of shots and templates. Fig. 5 shows an example of a tennis shot introduced by the user. To simulate the physical behavior of the virtual objects that compose the scene the Newton physics engine called Newton Game Dynamics [24] is used across an abstraction level that we define in [25], which allows a transparent communication between both software packages. The condition of every physic primitive is defined by its position and orientation in a certain time. The physical engine is the manager that calculates the new condition of the objects and updates it. The graphical engine requests this information for every object and applies the new values of position and orientation visualizing them on the screen.

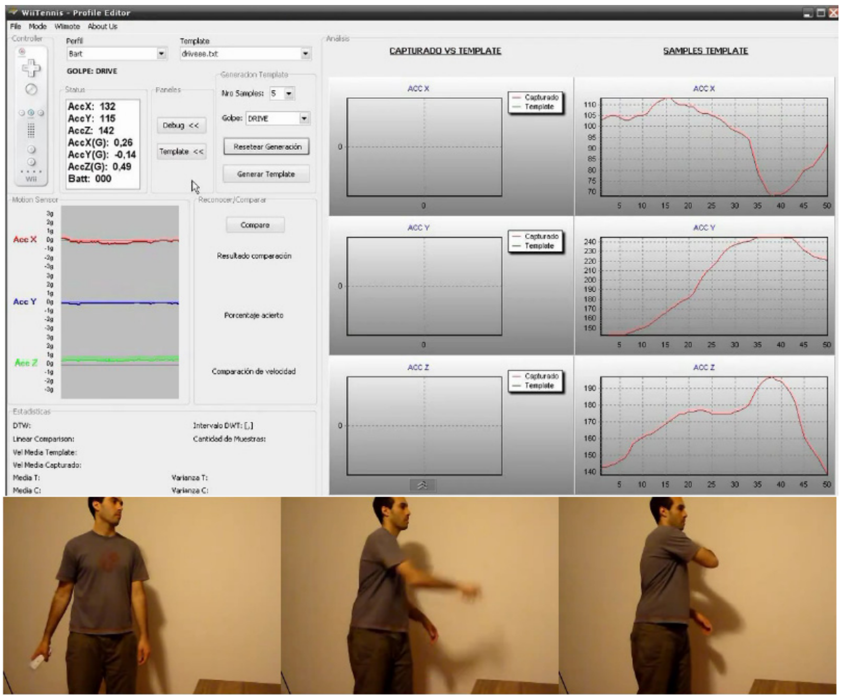

Fig. 5. Profile edition module: three video frames of the sequence of a tennis shot

4.2 Interaction

The application supports different modes of use and interaction: the training mode and the obstacles mode. In the training mode (Fig. 6) the user can execute different shots being able to check a series of statistical information as speed and place of rebound. It can compare his current training condition with a database of historical statistics. It can also compare his shot with a different user profile (for example one of a trainer or a professional tennis player). In the obstacles mode (Fig. 7) the application raises different fictitious situations with virtual objects to allow the training of directional shots. In these scenes the player must knock down objects in specific positions or succeed shots in different strategic points.

5 Experiments and Results

Our system is instantiated in a tennis training application that is called *WiiSimTennis.* To test the application we experiment with ten different users which already have experience in the use of the Wiimote. The experiment consists in the practice of five tennis shot: forehand, backhand, volley, smash and service [26].

The system was trained with 10 users to create a gestures database. The profile was also stored to uniquely identify each user. The first test consists in comparing a user shot with the gestures in his profile. This test evaluates the consistency of the algorithm.

Every user performed a series of 3 tests of every shot. For every shot the system calculates the measure of similarity with the gesture templates stored in the database.

Table 1 shows the results of this evaluation. The numbers indicate the percentage of similarity between the gesture that has fulfilled the user and the existing one in the database. As all the percentages are above 70 % we can assume that the system can recognize the shot without ambiguities. Also we can observe that the range of difference of 5 % in all the shots indicates us that the different users repeat the same shot in a similar way including the complex ones such as volleys and smash.

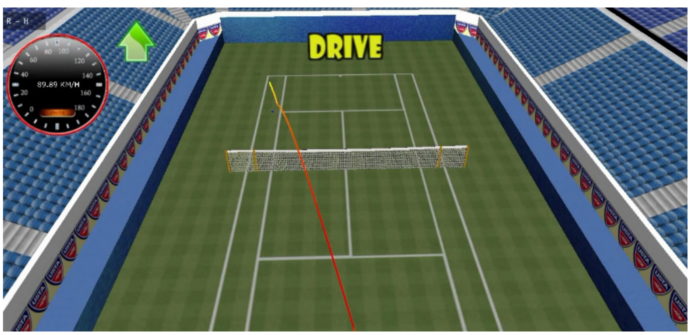

Fig. 6. Screen capture of the *WiiSimTennis* application: training mode

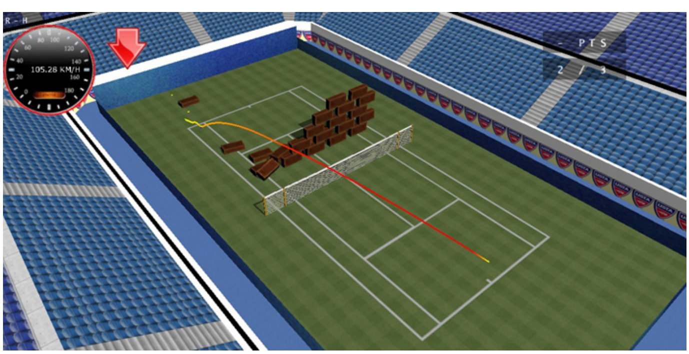

Fig. 7. Screen capture of the *WiiSimTennis* application: obstacles mode

Table 1. Percentage of similarity between the different tennis shots performed by ten different users with the existing template in the database

	Forehand	Backhand	Volley	Smash	Service
U1	84%	83%	78%	80%	81%
U2	81%	81%	81%	80%	81%
U3	83%	81%	82%	79%	80%
U4	83%	80%	79%	78%	78%
U5	82%	79%	80%	79%	80%
U6	84%	80%	81%	82%	80%
U7	85%	82%	79%	81%	79%
U8	83%	81%	78%	77%	81%
U9	82%	80%	79%	78%	82%
U10	80%	80%	79%	80%	81%
Mean	**83%**	**81%**	**80%**	**80%**	**80%**

We conducted a second test, where an experienced tennis player was chosen as reference. The remaining nine participants made each kind of tennis shot and the system tried to identify them. In Table 2 we show a confusion matrix with the obtained results

Table 2. Number of movements detected by the system according to the ones performed by the users

		Actual Class					
		Forehand	*Backhand*	*Volley*	*Smash*	*Service*	*unknown*
Predicted class	*Forehand*	**9**	0	0	0	0	0
	Backhand	0	**8**	1	0	0	0
	Volley	0	2	**7**	0	0	0
	Smash	0	0	0	**6**	2	1
	Service	0	0	0	2	**7**	0

The results seem acceptable. Some false positives that arise are "reasonable." For example, a *volley* and *backhand* are quite similar between them, so is the "service" and "smash". As shown in Table 1, the similarity is high in any situation and is true for both cases. When "false negatives" appear, the system applies a default movement.

Some other data can be extracted from the system. The percentage of similarity regarding a professional tennis player can indicate the level of tennis experience for each user. Additionally, these values can be analyzed to know the evolution of the user performance over the time.

6 Conclusions and Future Work

This paper presents a sport training system instantiated in a tennis training game called *WiiSimTennis*. The user interacts by means of the Wiimote in the 3D environment that represents the scene of the game and includes virtual object with Newton physic dynamics. The system can recognize the different user shots and presents information such as speed, place of bound and comparisons with historical information. Provided that the process of generation and validation of gestures is generic the system is applicable to a family of applications. Besides the tennis training application our system can be easily instantiated to other sports such as golf, bowling, basket and also a rehabilitation application. The instantiation implies the definition of the adapted 3D scene and the specific training of the application.

At present we are adjusting the system to be able to use a more complete input device such as the Microsoft Kinect.

As a future work we mention the inclusion of biomechanics to the system.

Acknowledgments. This research was partially supported by project A1/037910/11 FRIVIG. Formación de Recursos Humanos e Investigación en el Área de Visión por Computador e Informática Gráfica , granted by MAEC-AECID (Programa de Cooperación Interuniversitaria e Investigación Científica entre España e Iberoamérica).

References

1. Bowman, D.A., Kruijff, E., LaViola, J.J., Poupyrev, I.: 3D User Interfaces: Theory and Practice. Addison Wesley Longman Publishing Co., Inc., Redwood City (2004)
2. Perrin, S., Cassinelli, A., Ishikawa, M.: Gesture recognition using laser-based tracking system. In: FGR, pp. 541–546 (2004)
3. Turk, M.: Gesture recognition. Handbook of Virtual Environments, pp. 223–238 (2001)
4. Vafadar, M., Behrad, A.: Human Hand Gesture Recognition Using Motion Orientation Histogram for Interaction of Handicapped Persons with Computer. In: Elmoataz, A., Lezoray, O., Nouboud, F., Mammass, D. (eds.) ICISP 2008. LNCS, vol. 5099, pp. 378–385. Springer, Heidelberg (2008)
5. Caridakis, G., Karpouzis, K., Pateritsas, C., Drosopoulos, A., Stafylopatis, A., Kollias, S.: Hand trajectory-based gesture recognition using self-organizing feature maps and markov models. In: 2008 IEEE International Conference on Multimedia & Expo., ICME (2008)
6. Ruffaldi, E., Filippeschi, A., Avizzano, C.A., Bardy, B., Gopher, D., Bergamasco, M.: Feedback, Affordances and Accelerators for Training Sports in Virtual Environments. Presence, Teleoperations and virtual environments 20(1), 33–46 (2011)
7. Müller, H., Schumacher, B., Blischke, K., Daugs, R.: Optimierung sportmotorischen Technik-Trainings durch computergestutzte Videosysteme. In: Perl, J. (ed.) Sport und Informatik, pp. 37–47. Hofmann, Schorndorf (1990)
8. Schorer, J.: Eine Studie zur Identifikation, den Mechanismen und Entwicklung senso-motorischer Expertise. Dissertation. Universität Heidelberg (2006)

9. Savelsbergh, G.J.P., Williams, A.M., Van der Kamp, J., Ward, P.: Visual search, anticipation and expertise in soccer goalkeepers. Journal of Sports Sciences 20, 279–287 (2002)
10. Liu, X., Sun, J., He, Y., Liu, Y., Cao, L.: Overview of Virtual Reality Apply to Sports. JCIT: Journal of Convergence Information Technology 6(12), 1–7 (2011)
11. Bideau, B., Kulpa, R., Vignais, N., Brault, S., Multon, F., Craig, C.: Using Virtual Reality to Analyze Sports Performance. IEEE Computer Graphics and Applications 30(2), 14–21 (2010)
12. Ruffaldi, E., Filippeschi, A., Avizzano, C.A., Bardy, B., Gopher, D., Bergamasco, M.: Feedback, Affordances and Accelerators for Training Sports in Virtual Environments. Presence, Teleoperations and virtual environments 20(1), 33–46 (2011)
13. Christian, J., Krieger, H., Holzinger, A., Behringer, R.: Virtual and Mixed Reality Interfaces for e-Training: Examples of Applications in Light Aircraft Maintenance. In: Stephanidis, C. (ed.) HCI 2007, Part III. LNCS, vol. 4556, pp. 520–529. Springer, Heidelberg (2007)
14. Tanaka, K.: Virtual Training System using feedback for sport skill learning. Int. J. of Computer Science in Sport. 8(2), 1–7 (2009)
15. Siemon, A., Wegener, R., Bader, F., Hieber, T., Schmid, U.: Video Games can Improve Performance in Sports. An Empirical Study with Wii Sports Bowling. In: Wallhoff, F., Rigol, G. (Hrsg.) Proceedings of the KI 2009 Workshop on Human-Machine-Interaction (2009)
16. Chow, Y.: 3D Spatial Interaction with the Wii Remote for Head-Mounted Display Virtual Reality. World Academy of Science, Engineering and Technology 50, 377–383 (2009)
17. Leong, T., Lai, J., Panza, J., Pong, P., Hong, J.: Wii Want to Write: An Accelerometer Based Gesture Recognition System. In: International Conference on Recent and Emerging Advanced Technologies in Engineering, pp. 4–7. Carnegie Mellon University (2009)
18. Senin, P.: Dynamic time warping algorithm review. Technical Report CSDL-08-04, Department of Information and Computer Sciences. University of Hawaii, Honolulu, Hawaii 96822 (2008)
19. Niezen, G., Hancke, G.: Gesture recognition as ubiquitous input for mobile phones. In: International Workshop on Devices that Alter Perception (DAP 2008), Conjunction with Ubicomp (2008)
20. Liu, J., Wang, Z., Zhong, L., Wickramasuriya, J., Vasudevan, V.: uWave: Accelerometer based Personalized Gesture Recognition and Its Applications. In: IEEE PerCom (2009)
21. Wilson, D.H., Wilson, A.: Gesture Recognition using the XWand. Technical Report CMURI-TR-04-57. CMU Robotics Institute (2004)
22. Holt, G.A., Reinders, M.J.T., Hendriks, E.A.: Multi-Dimensional Dynamic Time Warping for Gesture Recognition. In: Thirteenth Annual Conference of the Advanced School for Computing and Imaging (2007)
23. D'amato, J.P., García Bauza, C.: Simulación de Escenarios Tridimensionales Dinámicos, Grade thesis dissertation. Universidad Nacional del Centro de la Pcia de Buenos Aires, Tandil, Argentina (2004)
24. Newton Game Dynamics - Physics Engine, http://newtondynamics.com/forum/newton.php
25. García Bauza, C., Lazo, M., Vénere, M.: Incorporación de comportamiento físico en motores gráficos. In: Mecánica Computacional, vol. XXVII, pp. 3023–3039 (2008)
26. Wikipedia, Tennis Shots, http://en.wikipedia.org/wiki/Tennis_shots