Chapter 16
Preservation and gamification of traditional sports

Yvain Tisserand, Nadia Magnenat-Thalmann, Luis Unzueta, Maria T. Linaza, Amin Ahmadi, Noel E. O'Connor, Nikolaos Zioulis, Dimitrios Zarpalas, Petros Daras

Abstract This chapter reviews an example of preservation and gamification scenario applied to traditional sports. In the first section we describe a preservation technique to capture intangible content. It includes character modelling, motion recording and animation processing. The second section is focused on the gamification aspect. It describes an interactive scenario integrated in a platform that includes a multi-modal capturing system, a motion comparison and analysis as well as a semantic based feedback system.

Keywords: cultural heritage, gamification, sport preservation, character modelling, motion recording, motion comparison

Y. Tisserand and N. Magnenat-Thalmann
MIRALab CUI, University of Geneva, Geneva, SWITZERLAND
e-mail: yvain.tisserand@miralab.ch; thalmann@miralab.ch

L. Unzueta, M. T. Linaza
Vicomtech IK4, San Sebastian, SPAIN
e-mail: lunzueta@vicomtech.org; mtlinaza@vicomtech.org

A. Ahmadi, N. E. O'Connor
INSIGHT, Dublin City University, Dublin, IRELAND
e-mail: amin.ahmadi@dcu.ie; noel.oconnor@dcu.ie

N. Zioulis, D. Zarpalas, P. Daras
Centre for Research and Technology Hellas, Information Technologies Institute, Thessaloniki, GREECE
e-mail: nzioulis@iti.gr; zarpalas@iti.gr; daras@iti.gr

16.1 INTRODUCTION

Traditional Sports and Games (TSG) are a strong part of the identity of a society and a strong mechanism for the promotion of cultural diversity. This chapter aims to explain research done in order preserve and promote these TSG. The rich intangible cultural heritage expressed through TSG is inherently gamified. The preservation and promotion of the cultural diversity through TSG is a challenging task due to trends in spectator and participative sports. Main stream and commercial sports are generating more interest due to their access to high technology and mass media outreach. By developing a technological platform around the interpretation of digital content for TSG through a popular modern medium like gaming, their reach to wider audiences and their access to the general public will be increased. This work has been achieved within the framework of the EU RePlay project.

This work can be divided into two main steps. First of all, in order to preserve the TSG, the original game has to be captured. We propose to capture, model and animate targeted TSG in 3D. The second step is to promote TSG, a gamification scenario will be described as well as it components. It includes an interaction system, a real-time multi-modal 3D capturing system, a motion comparison module and a semantic based feedback system.

16.2 GAMIFICATION FOR TRADITIONAL SPORTS AND GAMES

16.2.1 Platform overview

We present a multi-modal 3D capturing platform coupled to a motion comparison system, in the context of a PLAY&LEARN scenario, which is based on the definition of storylines which highlight its main features and are used to extrapolate from the present into the future of TSG. The TSG considered in this work are the Gaelic sports from Ireland and Basque sports from France and Spain.

Next, we show a couple of examples of this kind of storylines, and how the platform can be designed and built accordingly:

John is a 10 year old boy who plays Hurling at school. For his birthday, he got a Microsoft Kinect sensor with a new game called "Play against your Heroes". As

stated in the box, the game allows John to play against several players. However, John is only interested in Hero1, who is one of his favourite Hurling players. The game should be played along and also includes a gadget that looks much like a Hurley. When the game starts, John has first to choose among the existing players. Of course, he chooses Hero1. He knows Hero1 is an expert with the movement1, so he will try to mimic that movement1. The screen is divided into two parts. On the left part, John can see the movement1 played by Hero1. On the right part, he sees himself with the Hurley trying to mimic the movement. The Kinect captures his movement and presents it on a "puppet-like" avatar of himself. As he can see, there are several differences between his shot and movement1 from Hero1. Thus, he tries again to improve his performance.

Sarah is a 12 years old Pala player who wants to see how much she is improving her game. She has a Kinect at home, and she decides to buy the "Basque Ball game". On the screen, she can see herself and the National Hero or Heroine she has already chosen to compare her performance and the one of the National Heroine and see how far is from it. She tries to improve the movement, and she can check on the screen the trials she is doing and see if she is improving or not. She can play against any player too.

Thus, the PLAY&LEARN scenario focuses on children and teenagers having access to a low-cost motion capture set-up (e.g., one Kinect sensor or a set of Kinect and WIMU sensors) at home or school, to learn and mimic the skills of a National/Local Hero. The user can optionally have a copy of an instrument related to the selected TSG (e.g., a Hurley, a Cesta or a Pala). The main goal of this scenario is to promote the TSG to children and thus encourage their participation. Users can learn, compare and compete in the performance of sporting gestures and compare themselves to real athletes. Regarding the application, the player can initially configure his/her preferences (language, modality, the number of trials, hand to play). Then, the user must follow several steps (gain control of the Microsoft Kinect, watch the instructions and the 3D representation of the skill) before performing the skill. The user gets visual (two avatars side by side), semantic textual and score feedback.

A final issue is the estimation of the trajectory of the ball associated to the movement. As this is a home scenario, it does not seem to be feasible to have a real ball. However, the capturing platform can also estimate the ball trajectory to provide feedback to the player, if WIMU sensors are included in the setup. This feedback does not show the real place where the ball should be, but positive/negative feedback based on the estimated trajectory of the ball and the accuracy of the performance of the skill. This is important for a positive reinforcement in an engaging strategy.

16.2.2 Description of the infrastructure

In this scenario, different hardware configurations may be used. The simplest one consist of considering only one Microsoft Kinect sensor, while the most complex one considers a set of Kinects placed around the capture space and WIMU sensors placed on the user's body and the instrument related to the selected TSG. These different configurations depend on the space available for the setup and the desired level of precision for the capture. **Fig. 16.1** shows a setup where one Kinect sensor is used, and the user wears a set of WIMU sensors.

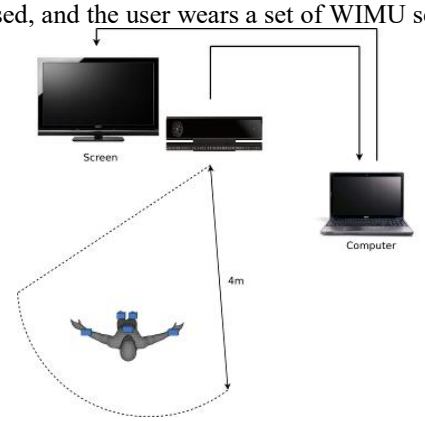


Fig. 16.1: *Layout of the platform for the PLAY&LEARN scenario.* ***(Vicomtech)***

In this specific setup, one Kinect sensor is used to capture the side-frontal view of the user. It is connected via USB to a PC running corresponding software for capturing and storing the streams. It should be placed no further than 4m from the user. Regarding the WIMU sensors, they need to be placed at different segments of a subject. The WIMU data can be transferred to the PC via Bluetooth connection.

Correct sensor placement will provide the best body tracking performance for the Microsoft Kinect device. The sensor needs to be placed in a location so that it can see the entire body of the performer.

The sensor should be positioned between 0.6m and 1.8m from the ground; ideally at least 15cm above digital screens; and also away from any speakers (at least 0.3m). Additionally, the sensor needs to be placed near the edges of flat surfaces; otherwise, its bottom view will be clipped. Another important issue is the lighting conditions in which the sensor operates. The surrounding space needs to have enough bright light and be equally lit. In any case, direct sunlight has to be avoided, thus it needs to be placed away from windows, or they should be shaded during daylight usage.

Finally, in cases of reflective floors, it is recommended to place the sensor 1.1-1.2 m above the ground parallel to the floor. Otherwise, it should be placed lower (0.8 m above the floor) and rotated so that the depth camera only captures the user without the ground (**Fig. 16.2**). Such rules are not restrictive as long as the sensor can see the entire body of the user and the user can freely move around without having obstacles limiting the view of the sensor.

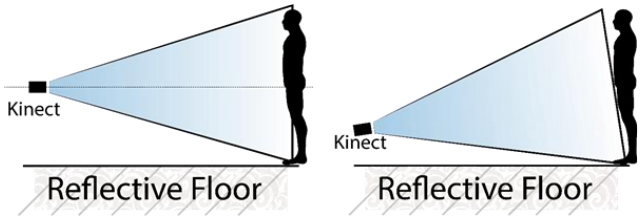


Fig. 16.2: *Placement of the Microsoft Kinect in case of reflective floors.* ***(CERTH)***

Besides the correct placement, extra attention has to be paid while using the Microsoft Kinect device. Despite being a state-of-the-art marker-less motion tracking sensor, it suffers from some limitations regarding self-occlusion that need to be taken into account during its usage. First, users should try to keep most of their body parts directly visible by the sensor. Secondly, as the sensor was designed for frontal usage mainly, with some rotational tolerance, it is recommended to keep the angle between the coronal plane of the user and the viewing direction of the sensor less than 45° (**Fig. 16.3**).

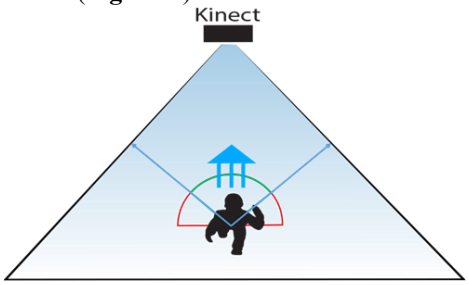


Fig. 16.3: *Placement of the Microsoft Kinect for correct use (top view).* ***(CERTH)***

The platform including the Kinect and WIMU sensors should be portable and easy to handle in a Plug&Play mode. However, if the full configuration is used a calibration phase must be considered, which should be done by a person with a basic technical knowledge about the system. Besides, the platform works on the basis of the "Quick Post" concept. This means that important feedback will be

given as soon as possible while other statistics should be provided later. In this way, the platform should give direct feedback.

16.2.3 Interaction experience of the user

Fig. 16.4 displays a diagram showing the relationship between the modules and components of the capturing platform.

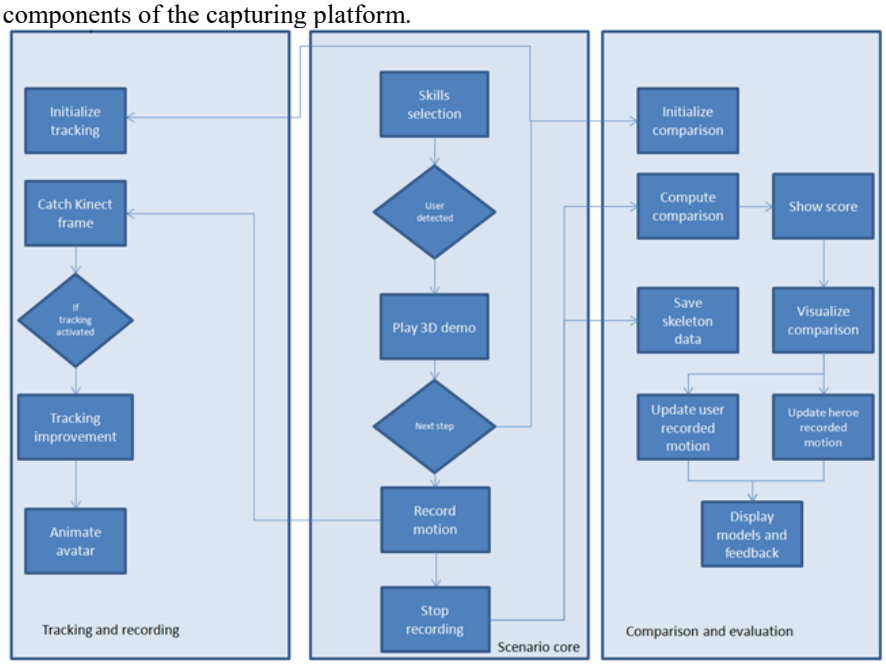


Fig. 16.4: *Combination of modules and components for the coach application of the PLAY&LEARN scenario.* ***(MIRALab)***

The user opens the application at the "Preferences" screen (**Fig. 16.5**) to select the language in which the application will run; the modality (Gaelic sports, Basque Pelota); the skills to be played; the number of trials; the hand normally used to play; and whether he/she is going to use the application as an experienced or non-experienced player.

Fig. 16.5: *Selection of the preferences in the PLAY&LEARN scenario.* ***(Vicomtech)***

Fig. 16.6 displays the workflow of the capturing platform. To start the trial, the user must wave his/her right hand to be recognized by the Kinect sensor and to gain control of the movements of the avatar. Then, the user can watch the instructions to perform the current skill. Afterwards, the user can watch a 3D representation of the skill performed by a National Hero. Finally, the user can perform the skill after the countdown, when the "Go" alert appears. In order to compare the skill, two avatars are presented side by side, one representing the performance of the National Hero and the other one representing the player (**Fig. 16.7**).

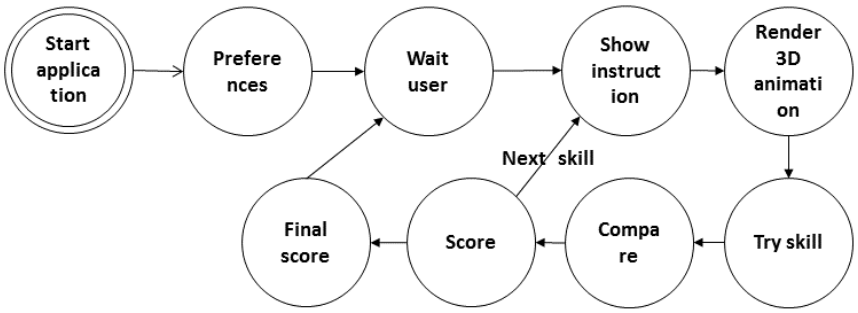


Fig. 16.6: *Global flow of the PLAY&LEARN application.* ***(MIRALab)***

Fig. 16.7: *Several screenshots of the capturing platform for the PLAY&LEARN scenario* ***(MIRALab)***

16.3 TRADITIONAL SPORT AND GAME CAPTURE, MODELLING AND ANIMATION

To capture TSG skills in 3D, three main steps are required. First of all, we need to create the shape of the 3D avatar that represents the athlete, then we need to capture its movement and, finally, we need to animate the 3D avatar according to the captured animations.

16.3.1 Avatar creation

The time-consuming manual process of avatar creation has been replaced over time by several techniques. Different methodologies have been proposed and can be classified into three mains categories: creative (Ratner, 2012), reconstructive (Allen et al., 2003) and interpolation methods (Bastioni et al., 2008). We propose a reconstruction based technique that uses an image-based 3D scanner to capture the user in a fast and accurate manner. The post-processing time and the cost of the installation can be significantly decreased, compare to the previous generation of body scanner, such as laser-based body scanner.

The system is based on photogrammetry technologies. It is composed of a large number of compact cameras that are synchronized and controlled by a computer. Within less than a second, pictures of the subject are taken by the camera cluster from different angles. This very short delay during the capture minimizes user movements, which drastically reduces the noise in the generated model. The images can then be used for 3D reconstruction. Finally, a virtual skeleton is inserted into the model to be able to animate it.

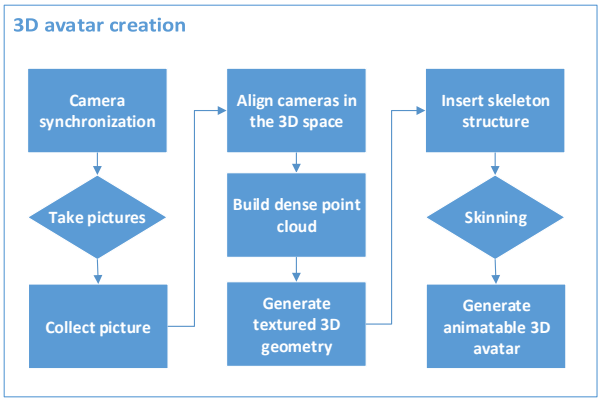


Fig. 16.8: *3D avatar pipeline* ***(MIRALab)***

Our system is composed of a cluster of 80 compact cameras. They have been placed onto a hexagonal structure. Our acquisition volume covers an adult human, and the number of cameras and their positions have been chosen accordingly. A made-to-measure green fabric has been placed over the support structure to control the light conditions and to facilitate the post-processing of the acquired data. To get diffuse light inside the scanner, flexible led ribbons have been attached to the support structure.

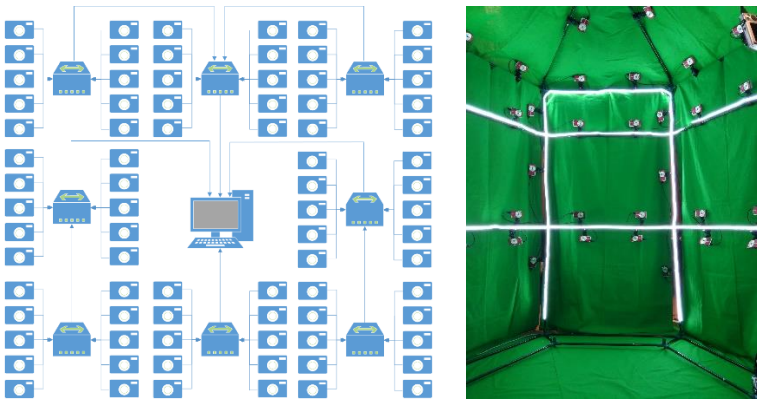


Fig. 16.9: *Overview of the image-based 3D scanner* ***(MIRALab)***

All cameras are connected to a single computer. A dedicated library has been used to control, to synchronize and to take pictures with the camera remotely (CHDK). Custom scripts have been written to remotely control and synchronize the individual cameras, to adjust the zoom, to take shots and to copy back recorded pictures to the controller computer. After a short synchronization step, we can remotely take a synchronized shot.

The user simply has to stand inside the structure and to hold the position for a second. Once the pictures are taken, they are automatically copied to the hard drive of the controller computer for reconstruction.

We use an image-based 3D reconstruction software (Agisoft PhotoScan) to generate the 3D avatar mesh. As input, it requires a set of images. An optional mask can be used for accelerated 3D reconstruction. The process can be divided into four steps: camera alignment, point cloud creation, mesh reconstruction and texturing (see **Fig. 16.10**).

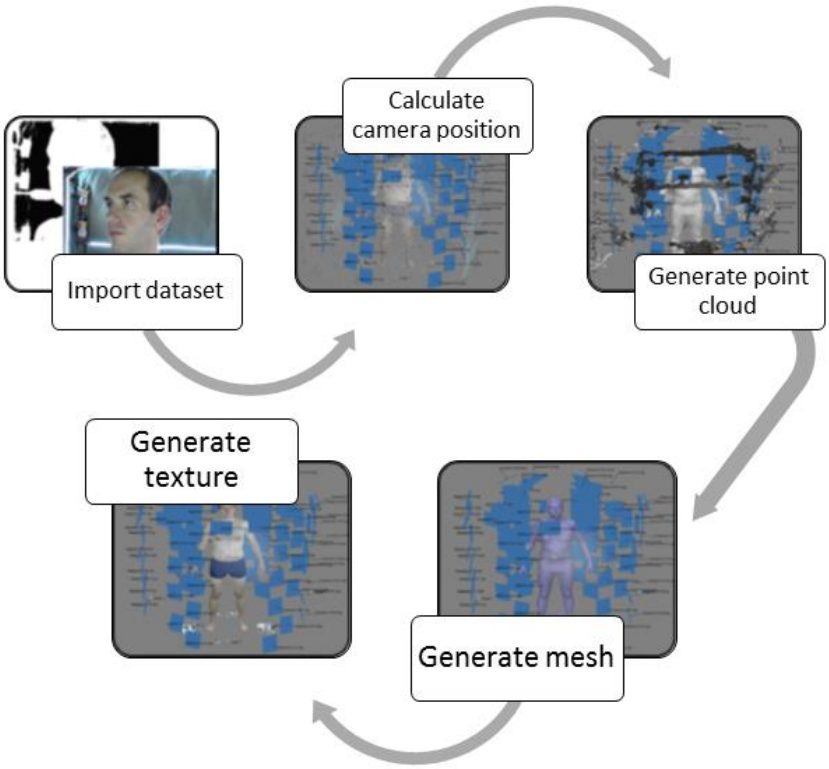
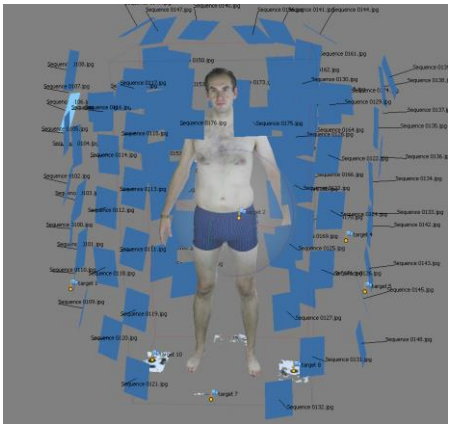


Fig. 16.10: *3D reconstruction pipeline using image-based method* ***(MIRALab)***

The camera alignment consists of two sub-steps. First, features are detected in all images. In a second step, the software tries to match the features pair-wise in the set of images. Therefore, a sufficient overlap of the images is needed. It can be achieved by carefully controlling position and zoom level of the cameras. Several tests have been conducted to develop our current setup.



a. Textured model

b. Cleaned model

Fig. 16.11: *Example of 3D reconstructed avatar* ***(MIRALab)***

We obtain a fully reconstructed and textured 3D. However, small corrections are needed to remove mesh artefacts in the obtained 3D model (**Fig. 16.11**). First, we apply Laplacian smoothing to reduce the noise and to smooth the mesh. Then, to reduce the number of polygons and to get a regular grid on the 3D mesh, we apply a Quadric Edge Collapse Decimation algorithm. As results, we obtain a static mesh that represents the athlete.

a. Dressed avatar

b. Animated avatar

Fig. 16.12: *Rigging, clothing and animation of the 3D avatar* ***(MIRALab)***

A final step is then required in order to have a fully animatable 3D avatar. A virtual skeleton has to be added as well as virtual garments (**Fig. 16.12**). The result

is a fully functional dressed 3D avatar that can be used and animated in any 3D platform.

16.3.2 Motion capture

To capture sports skills with a high level of precision, we have chosen an optical motion capture system provided and controlled by Vicon. Due to sports constraints such as skill's speed, field dimension or the specificity of accessories; a particular setup has been defined. A motion capture studio has been used with a large tracking space volume.

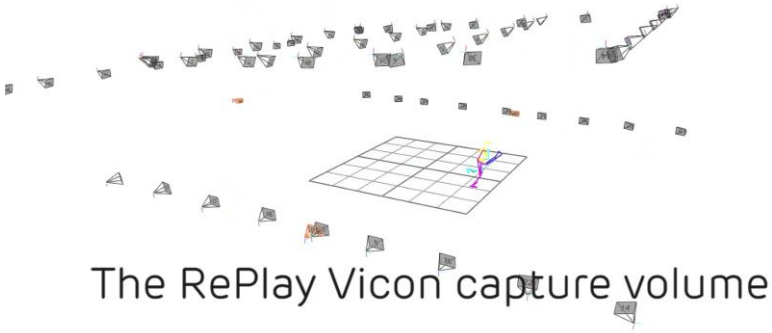


Fig. 16.13: *Tracking camera setup* ***(VICON)***

50+ cameras have been used to track the athletes' movements with a high accuracy (**Fig. 16.13**). The "PlugIn Gait" marker setup has been used. In is composed of 45 reflective markers used to track the full human body (**Fig. 16.14**). Some extra markers have also been used to track sports accessories.

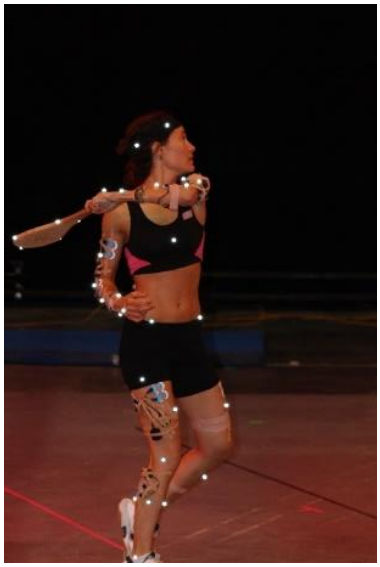


Fig. 16.14: *Athlete being captured* ***(MIRALab)***

As results, we obtained 180 records from 8 athletes of different TSG. The output is an animation file that can be combined with a virtual avatar or that can be applied to a stick figure.

16.3.3 Avatar animation

To animate the 3D avatar, a skeleton structure has to be added to the static mesh made using the image-based 3D scanner. The avatar is then animated by mixing the 3D avatar obtained using the 3D scanner and kinematic data. The output is a high-quality animation that includes a high frame rate (200fps) and high level of precision. It is possible to render the animation with all recent graphical engine, including Unity3D and Unreal Engine. To increase the immersion, we designed some 3D environment to reproduce the TSG fields.

Fig. 16.15: *3D avatar animated and placed in the 3D environment* ***(MIRALab)***

The animated avatar is then placed in a virtual context and can be shown to promote and preserve in a gamification context (**Fig. 16.15**).

16.4 REAL-TIME TRACKING

We chose wearable inertial sensors and the Microsoft Kinect since they are low- cost and are each gaining in popularity in the area of human movement monitoring and gesture recognition due to their accuracy and potential for real-time applications. In the following section, we introduce these two sensor modalities as well as describing the advantages and disadvantages of each sensor with respect to motion capture.

Kinect: Since very recently, computer game users can enjoy a novel gaming experience with the Xbox, thanks to the introduction of the Microsoft Kinect sensor, where your body is the controller. Like the Nintendo Wii sensor bar, the Kinect device is placed either above or below the video screen. However, the Kinect adds the capabilities of a depth sensor to those of an RGB camera, recording the distance from all objects that lie in front of it. The depth information is then processed by a software engine that extracts, in real time, the human body features of players, thus enabling the interaction between the physical world and the virtual one. However, there are some disadvantages associated with Kinect including low

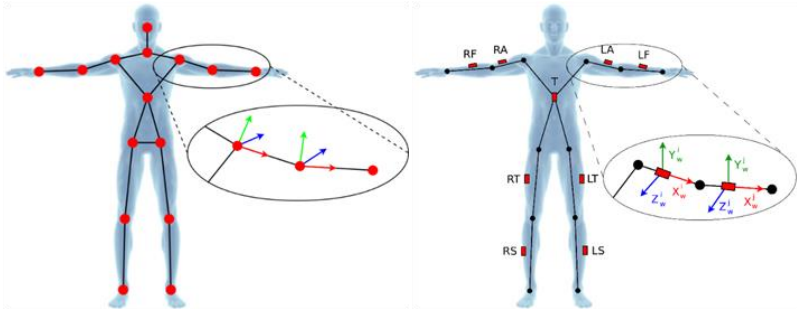
frame rate, limited volume of capture, inaccurate joint orientation estimation and lighting and occlusion problems.

WIMU: In general, a Wireless/Wearable Inertial Measurement Unit, or WIMU, is an electronic device consisting of a microprocessor board, on-board accelerometers, gyroscopes and magnetometers and a wireless connection to transfer the captured data to a receiving client. WIMUs are capable of measuring linear acceleration, angular velocity, and gravitational forces and are often used in MoCap systems. MEMS inertial sensors are being widely used in MoCap research due to the following reasons:

- They are miniaturized and lightweight so they can be placed on any part or segment of a human body without hindering performance.
- The cost of such sensors is falling dramatically as they start to persuade mass market consumer devices.
- They can be utilized to capture human movement/actions in real unconstrained environments (e.g. outdoor environments with variable lighting conditions) to obtain accurate results.
- They can be used to provide real time or near real time feedback.

A possible solution to the limitations of the Kinect system is to combine the Kinect based data with data from wireless inertial motion units (WIMUs) which can provide greater accuracy in the measurement of body segment angles and angular velocities, and also have much higher sampling frequencies (e.g. up to 1024 Hz) at consistent rates. WIMUs can incorporate tri-axial accelerometers and gyroscopes, to determine angular measures and facilitate an accurate identification of key events which involve impact (e.g. ground contact when jumping, striking a ball in tennis). The use of WIMUs alone, however, is limited because of significant challenges in determining accurate joint center position necessary in the provision of visual feedback on the body's motion. This provides the motivation for fusing information from Microsoft Kinect and multiple WIMUs.

Different capturing modalities used within the RePlay platform provided different types of skeletons. Using the Microsoft Kinect can generate a relatively simple skeleton with 16 bones and 15 nodes as shown Figure 1a. It also provides 3D segment angles linked to each bone. Using the Kinect and different number of WIMUs can result in generating a fused skeleton which is more robust, reliable and accurate than the skeleton generated by Kinect (**Fig. 16.16** b). The fused skeleton is the primary method to capture athletes' performance to be compared against that of a national hero within the RePlay platform. The main challenge to implement the fused skeleton is to obtain accurate 3D orientation using the WIUMs. In the following section, 3D orientation estimation using WIMUs is briefly discussed.



(a) (b)

Fig. 16.16: *(a) The 3D skeleton captured using the Microsoft Kinect sensor; (b) The 3D skeleton generated using Microsoft Kinect and 9 WIMUs are shown* ***(INSIGHT)***

16.3.5 Orientation Estimation using Inertial Sensors

Measuring accurate orientation plays an important role in sports activity applications as it enables coaches, biomechanists, and sports scientists to monitor and investigate athletes' movement technique in indoor and outdoor environments. Although there are different technologies to monitor athletes' technique and measure their body orientation, wearable inertial sensors have the advantage of being self-contained in a way that measurement is independent of motion, environment, and location. It is feasible to measure accurate orientation in 3D space by utilizing tri-axial accelerometers, gyroscopes, and a proper filter. We employed a filter that utilizes a quaternion representation, allowing accelerometer data to be used in an analytically derived and optimized gradient descent algorithm to compute the direction of the gyroscope measurement error as a quaternion derivative (Madgwick et.al., 2011, Ahmadi & Mitchell., 2015). The filter has been shown to provide effective performance at low-computational expense. Using such a technique, it is feasible to have a lightweight, inexpensive system capable of functioning over an extended period of time (Madgwick et.al 2011, Ahmadi & Mitchell., 2015).

16.4.1 Sensor placement

Each inertial sensor device (WIMU) has to be placed on one segment of a subject in a predefined orientation. The location of the sensor on each body segment was chosen to avoid large muscles; as soft tissue deformations due to muscle contractions and foot-ground impacts may negatively affect the accuracy of joint orientation estimates. As it is shown in **Fig. 16.17**, the sensors the X-axis and Y-axis of each sensor are well aligned with the longitudinal axis of the corresponding bone for the upper body and lower body segments, respectively. It should be noted that the number of sensors used to be fused with the MS Kinect sensor is selectable (between 1 to 9).

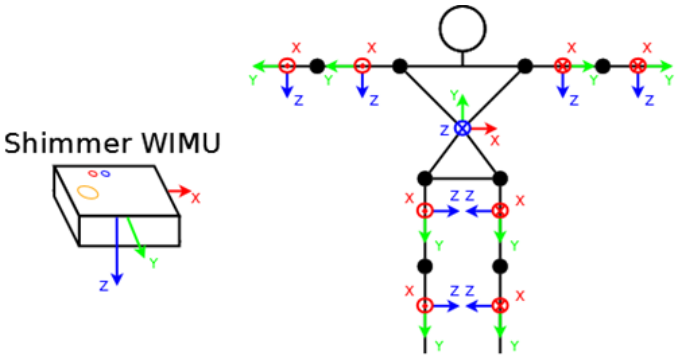


Fig. 16.17: *Shimmer sensor orientation (left) and the sensor placement on different segments of a subject is illustrated (right). Nine inertial sensors are fixed to the subject's forearms, arms, thighs, shanks and to the chest. These correspond respectively to the fused skeleton joints R/LF, R/LA, R/LT, R/LT and T.* ***(INSIGHT)***

16.4.2 Methodology

We designed and implemented the fused Kinect / WIMUs skeleton using three separate information sources given by each modality (Destelle et al., 2014). The Kinect sensor provides the initial joint positions of our skeleton, as well as the global position of the subject's body over time. The WIMUs provide the orientation information, which we need to animate each bone of our fused skeleton over time.

First, we consider a reference skeleton provided by the Kinect sensor and the associated skeleton extraction algorithm. This reference skeleton is the starting point of our fused skeleton synthesis method and is built from a reference frame captured by the Kinect. We need this reference skeleton to be as accurate as possible, to produces a stable result. In order to do this, subject is asked to stand still in a T-pose with his/her palms facing the ground in front of the Microsoft Kinect sensor for five seconds to successfully obtain the reference skeleton. This is shown in following **Fig. 16.18**.

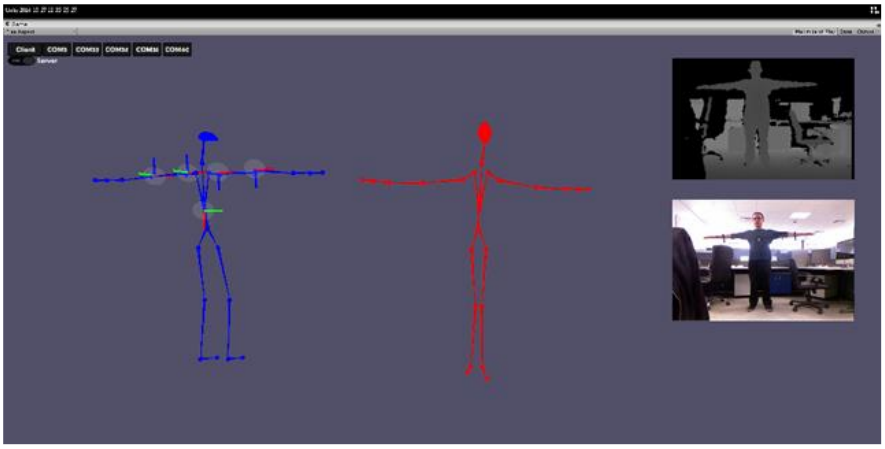


Fig. 16.18: *T-Pose required by the RePlay platform to calibrate the fusion of a Kinect sensor and the WIMUs* ***(INSIGHT)***

Secondly, for each subsequent frame captured by the two sensory modalities, we consider one specific joint captured by the Kinect algorithm, and the rotational data provided by the WIMUs. The aim of this specific Kinect skeleton joint is to track the global displacement of the subject's body over time, as the WIMUs cannot provide this information easily. For stability and simplicity purposes, we choose to consider the *chest/torso* joint of the Kinect skeleton. As a result, the location of the central joint of our fused skeleton is updated with respect to the displacement of this Kinect joint.

Finally, our fused skeleton is built from the reference skeleton. For each dataset captured by the WIMUs, each bone of our fused skeleton is rotated according to this rotational information in a hierarchical manner. This first process defines a new position for the starting and the ending points of our fused skeleton. For instance, the wrist position of the subject is affected by the orientation of the elbow, shoulder and torso, respectively. As such, once the orientation of the chest is obtained (as the root of the animated skeleton) then the remaining joint positions and orientations can be estimated. In other words, the rotated shoulder joints can be

used to calculate the new position of the elbow joints. The hip joints, in turn, can be used to calculate the new position of the knee joints.

Two examples of the accuracy of the fused skeleton (in blue) and the Kinect skeleton (in red) are shown in **Fig. 16.19**. The point clouds captured by the Kinect are superimposed. It can be seen that the skeletons captured using the Kinect are not as stable, reliable and accurate as those generated using the fused skeleton. For instance, it is evident in **Fig. 16.19** (left) that due to the occlusion issue, the left knee is not detected correctly by the Kinect. Also in **Fig. 16.19** (right), the right elbow, shoulder and knee joints were not detected correctly by Kinect since the subject was performing fast movements inside the capturing volume.

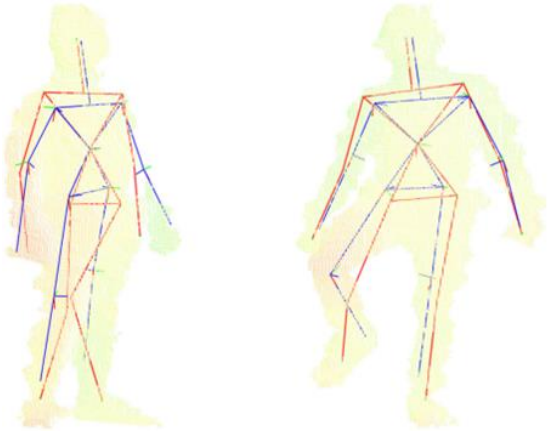


Fig. 16.19: *Two examples of the accuracy of the fused skeleton (in blue) versus the Kinect skeleton (in red).* ***(INSIGHT)***

Table 15.1 shows a comparison of the performance of the Microsoft Kinect and the fused skeleton compared to the VICON system. These results were measured during the performance of knee flexion, where the subjects were asked to raise their right knees up to about 90 degrees before flexing/extending their knees. The subjects were also asked to be in a predefined pose (T-pose) and then flex and extend their right and left elbows.

The root means square errors (RMSE) and the normalized cross-correlation coefficients (NCC) were measured during the whole trial (Destelle et al., 2014) . It can be seen that the results obtained from the fused skeleton is always closer to those measured using the VICON system. Five subjects participated in this experiment.

Joint angle	Left knee flexion RMSE	Left knee flexion NCC	Right knee flexion RMSE	Right knee flexion NCC
Kinect L-Elbow Fusion L-Elbow	16.73 14.19	0.13 0.70	9.93 3.81	0.61 0.85
Kinect R-Elbow Fusion R-Elbow	12.06 6.97	0.41 0.89	10.34 5.12	0.56 0.84
Kinect R-Knee Fusion R-Knee	29.51 6.79	-0.63 0.73	26.94 8.98	-0.02 0.50
Kinect R-Knee Fusion R-Knee	9.82 4.10	0.82 0.99	12.96 5.86	0.80 0.99

Table 15.1*: Numerical analysis of the Microsoft Kinect and the fused skeleton results compared to the results from the VICON system during the right knee flexion gesture with five trials* ***(INSIGHT)***

16.5 COMPARISON & FEEDBACK

Even though TSG are games by definition, the adhered to gamification approach in the context of a computer game aims to familiarize users with TSG in engaging and educative ways. Thus, two gamification elements were implemented, **i)** introduce digital game intrinsic principles like scoring that allow for competition among multiple players, and **ii)** combine it with an educational aspect to engage the users and facilitate their continuous skill improvement. The followed gamification approach relies on guiding the player's performance to match their favorite hero's one.

16.5.1 Compare and Score

Comparing a sports skill performance against a reference one, and produce a representing score poses as a big challenge, not only due to the complexity of the human motion, but also because of the following issues:

- Non-uniform representations: Different modalities are used to capture the performance of a professional hero and those of the players. Excluding the purpose of preserving national hero performances, they also need to be captured in very high quality and detail, to serve as the "reference" motions in our evaluation method and to drive the comparison results. To ensure this, the "gold standard" professional motion capturing system of VICON was used, while for the player skill performance motion capturing, the solution described in Section **d.** was used.
- Noisy measurements: Using low-cost sensors for the performance capture of the players can introduce varying levels of noise.
- Varying execution speed and,
- Coordination differences: Actual player performance deviates from that of the professional athlete due to inexperience and the skill level gap.
- Analysis of the motion: Each sport skill needs to be analyzed to identify its important characteristics and how its performance level can be assessed.
- Meaningful scoring and feedback: The outcome of the sport skill evaluation methodology should guide the player's improvement in performing it.

To overcome the above, the proposed solution's overview is illustrated in the logical pipeline of **Fig. 16.20**.

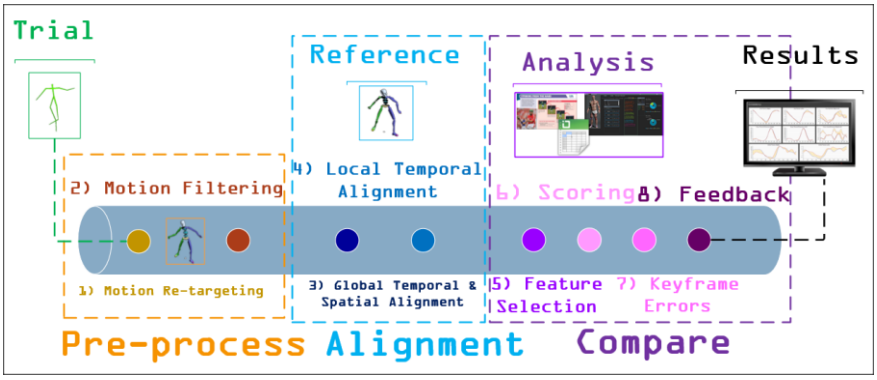


Fig. 16.20: *The motion evaluation pipeline* ***(CERTH)***

16.5.1.1 Pre-processing

The user's ("trial") captured motion has different signal characteristics than the captured hero performances ("reference"). As a result the first step of our pre-processing pipeline was motion re-targeting, where given a motion signal in a specific body structure format, it is transformed to another body structure format. Motion re-targeting is achieved by extending the methodology described in Ahmadi et al. 2015 only for the leg, to the whole body, so that the raw trial motion data captured by the low-cost motion capturing system are parameterized and re-

targeted to the reference motion's body structure. This step effectively offers the two motion sequences under comparison in a unified body format. Next, motion filtering is employed as the trial motion captured data contain noisy measurements that need to be filtered in order to correct erroneous pose estimations. Due to the nature of athletic action motion signals, characterized by sharp joint movements, an amplitude preserving filtering solution was opted for. More specifically, the least squares polynomial fitting Savitzy-Golay (Savitzy et.al. 1964) filters were selected, whose properties rather than being defined in the frequency domain and then translated to the time domain, they are derived directly from a particular formulation in the time domain aiming to preserve higher moments, while smoothing and supporting inflection at the same time. Another advantage of this choice is that due to its polynomial form the filter itself can be differentiated and, thus, the derivatives can be seamlessly calculated, providing the velocity and acceleration feature estimations.

16.5.1.2 Alignment

Having the trial and reference motions expressed uniformly in the same body structure enables the alignment phase. However, in order to be compared they need to be **i)** spatially and **ii)** temporally aligned. The two-step temporal alignment procedure initially estimates the global temporal offset and the spatial relative pose transformation between the two sequences, and then estimates local temporal alignment correspondences. In particular, quaternionic signal processing techniques are used by embedding the joint positions in pure quaternions and then first estimating their relative shift in time (global temporal alignment) through the maximum of the quaternionic cross-covariance similar to Alexiadis et.al. 2014. Secondly, local temporal warping through the Dynamic Time Warping technique is employed using the same quaternionic representation of the joint positions with respect to the pelvis joint, for the resulting globally aligned motions. The distance used for calculating the DTW path when using pure quaternion is the three-dimensional Euclidean distance. As a result the rotational invariance achieved through the global rotation between the two motions (encoded in the phase of their cross-covariance for time equal to zero) is instrumental to the local alignment step. This two-pronged temporal alignment strategy accounts for all the temporal inconsistencies either global (different start and end times as well as durations) or local (varying durations of each phase). Concluding, it should be noted that in the work of Alexiadis et.al. 2014 all the joints participated into the alignment calculations, but for the developed game, due to inexperienced users and in order to maintain high user experience levels while playing, the selection for the local alignment step was limited to one joint, the most informative one based on the weights defined in the following subsection. A visual example of the alignment methodology is shown in **Fig. 16.21**, with the "reference" motion in yellow and the "trial" in orange.

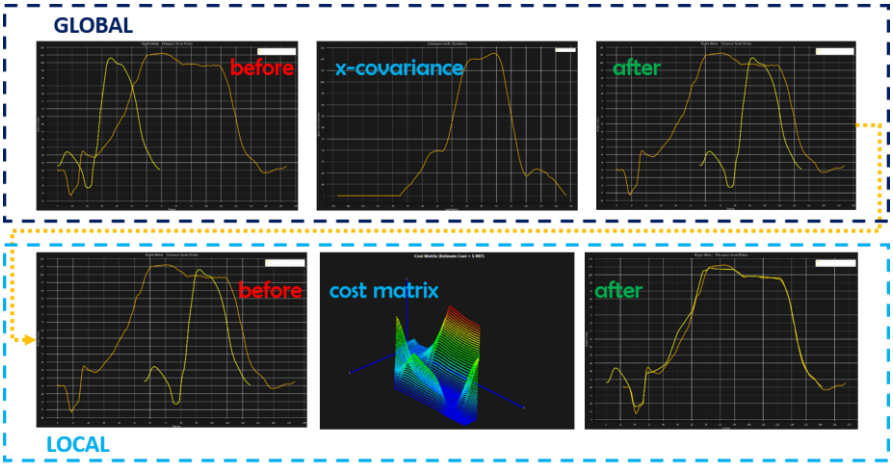


Fig. 16.21: *The motion alignment methodology. Top left: Right-wrist relative position to the pelvis – initial reference and trial motion features; Top Middle: The cross-covariance of the reference and trial features; Top Right: Trial and reference motion features after removing their global time shift; Bottom Left: The now temporally and spatially aligned motion features are fed to the local alignment algorithm; Bottom Middle: The 3D heat-map plot of the DTW cost matrix; Bottom Right: The final temporally aligned reference and trial motion features.* ***(CERTH)***

16.5.1.3 Compare

As aforementioned, the implemented evaluation scheme is a blend of achievement and incentive driven gamification principles with educative and learning elements. The learning element is based on offering teaching points, tailored to each specific skill, to guide players on how to correct their performance. These teaching points were mapped to motion features and a weighted scheme is utilized based on a selected subset of these features, corresponding to the teaching points, and their weighted relative importance to drive a hybrid comparison method.
More specifically, each sport skill was initially analyzed after taking into account its teaching instructions into a set of phases and features with respect to each phase. These phases are the **a)** Backswing, **b)** Frontswing and **c)** Follow through and are delimited by a set of key-frames: **i)** start of backswing, **ii)** start of frontswing, **iii)** ball impact and **iv)** end of follow through. The feature pool includes features like joint's velocity, acceleration and anthropometric angles (flexion, extension, abduction, adduction etc.)

Given two aligned motion sequences, each phase's features are extracted and compared using the Structural Similarity Index metric (SSIM) in order to offer a feature specific score. The SSIM proposed by Wang et.al 2004 in the context of image quality analysis is utilized after adapting it for use with one-dimensional time-series data instead of images and it is comprised of:

1. An amplitude term, scoring the average value of a set of measurements.
2. A measurement distribution term, scoring the variance of a set of measurements.
3. A structural term, scoring the temporal interdependencies of a set of measurements.

A weighted combination of these three terms is calculated for each feature's time-instant around a local neighborhood (in time) and then averaged for that feature's phase duration, calculating its score. Then, the overall score is computed by averaging all the features dictated by the motion analysis schema created for that skill.
An example is presented in the **Fig. 16.22**, where the performance of a *Handball – Right-handed Volley* skill is being assessed. The right elbow's flexion and the right hip's adduction were defined as two of the important motion features for this skill.

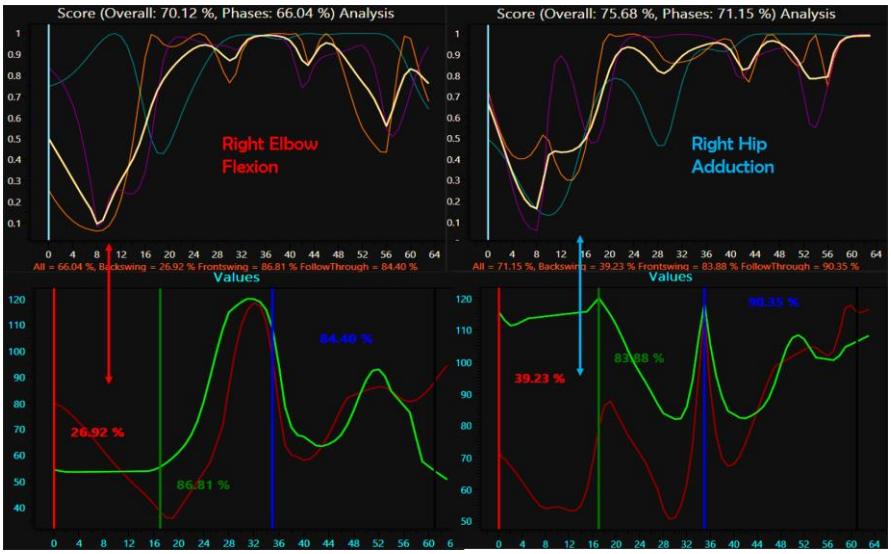


Fig. 16.22: *Per feature scoring analysis example. The trial (green) and reference (red) motion features at the bottom row and the SSIM and its respective terms at the top one (amplitude term in dark green, distribution term in orange, structural term in purple and overall SSIM score in light brown). The vertical lines represent the motion's key-frames (red for start of the backswing, green for start of the frontswing, blue for the impact point and black for the end of the follow through. The colored percentages denote each phase's score for that feature (red for the backswing, green for the frontswing and blue for the follow through).* ***(CERTH)***

16.5.2 Feedback and Visualization

This analysis and interpretation of the performed skill with respect to the defined teaching points is ultimately driving the educational aspect of the game, the offered feedback that players receive. This returned feedback decomposes the scoring overview and identifies specific sources of error and areas of improvement with respect to the important key-frames of the motion. The motion analysis schema associates specific features at specific key-frames of the motion with the required teaching points and semantic instructions around them. Then an error metric is calculated for each of these key-frame features that is then used to decide which instruction is to be triggered. Consequently, the hybrid comparison method uses both temporal technique scoring and key-frame posture error estimations to provide feedback in numerous ways:

- Score percentage.
- Semantic text feedback.
- Visual animation feedback.

First, the score is presented to the user with, optionally, detailed per feature scores and plots. Then two avatars are animated side by side using the alignment information to visually highlight corresponding postures during the performed action (**Fig. 16.23**). In parallel, this visual animation playback slows down when reaching erroneous key-frames, pauses and annotates, by color-highlighting, the body segments involved with the error and then displays the semantic feedback instructions to guide the player's improvement.

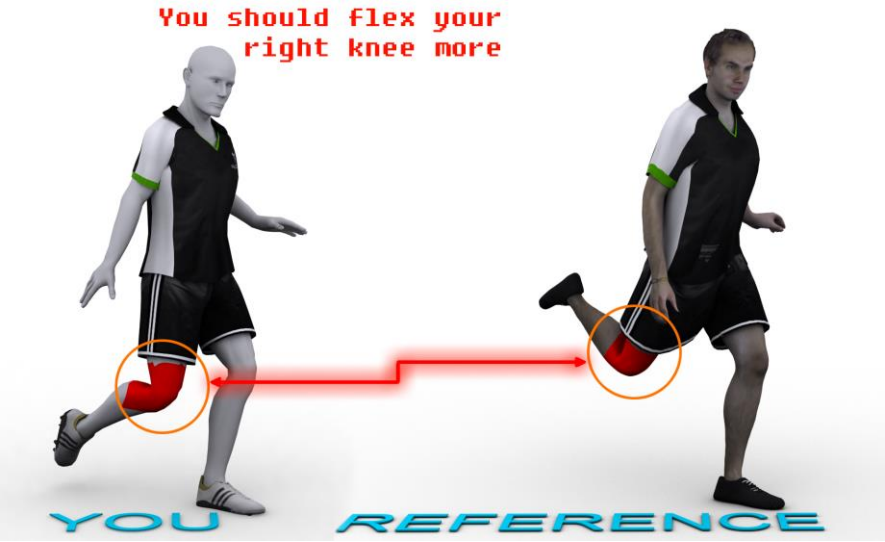


Fig. 16.23: *Visually annotated semantic feedback* ***(CERTH)***

16.6 CONCLUSION & RESULTS

The proposed platform has been implemented and tested together with TSG federation. A demonstrator has been setup for several events around Europe to promote TSG as well as the technology, where the participants filled a questionnaire about it. Thanks to the provided feedback, it has been concluded that one important parameter related to the ease of use is the understanding of the skill the participant had to mimic. In this case, the skill was completely understood by half of the participants and the remaining understood the skill quite well. Another important parameter related to the ease of use is the understanding of the comparison between the user and the National Hero to evaluate the visual feedback provided. The results of the questionnaires demonstrate that two thirds of the participants understood quite well the comparison, while the percentage is similar in the case of "A bit" and "Nothing". Regarding the expected results, half of the participants agreed that the score was quite approximate to what they would have expected. Regarding the understanding of the numbers in the score (general score and percentages for each of the phases), two-thirds of the participants agreed on having understood the numbers very well. One important feature of the feedback is the semantic feedback, providing text-based instructions at the bottom of the screen with suggestions to improve the score. Half of the participants read most of the instructions and one third read all the instructions. Related to this question, participants were

asked if they have followed the recommendations provided by the platform. The distribution of the responses clearly demonstrates that the feedback was taken into account mostly by half of the participants and completely by one third of them. The biggest issue regarding the demonstrator is the accuracy of the Microsoft Kinect sensor in capturing challenging sport actions. The discrepancy between the capture quality between the Gaelic field trials and the Basque ones resides in the increased difficulty of the Microsoft Kinect sensor to appropriately track a kicking action. While the capture quality of the Fist Pass skills at the same level of capturing quality as similar hand action skills, the capturing quality of the Punt Kick skill than half of that. Finally, due to either lighting conditions or sensor problems, the captured frame rates sometimes varied as seen in the frame rate distribution per trial. This issue can be improved in the future with the development of more robust techniques for human body pose estimation techniques from depth-sensing cameras, specifically designed for capturing motions of the targeted sports and games.

REFERENCES

[1] P. Ratner, 3-D Human Modeling and Animation, Wiley, 2012.

[2] Allen, B., Curless, B., & Popović, Z. (2003, July). The space of human body shapes: reconstruction and parameterization from range scans. In ACM Transactions on Graphics (TOG) (Vol. 22, No. 3, pp. 587-594). ACM.

[3] M. Bastioni, “Ideas and methods for modeling 3D human figures The principal algorithms used by MakeHuman and their implementation in a new approach to parametric modeling,” pp. 1–6.

[4] S. O. Madgwick, A. J. Harrison, and R. Vaidyanathan, “Estimation of imu and marg orientation using a gradient descent algorithm,” in *Rehabilitation Robotics (ICORR), 2011 IEEE International Conference, on*. IEEE, 2011, pp. 1–7.

[5] Ahmadi, A., Mitchell, E., Richter, C., Destelle, F., Gowing, M., O'Connor, N. E., & Moran, K. (2015). Toward automatic activity classification and movement assessment during a sports training session. *Internet of Things Journal, IEEE*, *2*(1), 23-32.

[6] Destelle, F., Ahmadi, A., O'Connor, N. E., Moran, K., Chatzitofis, A., Zarpalas, D., & Daras, P. (2014, September). Low-cost accurate skeleton tracking based on fusion of Kinect and wearable inertial sensors. In *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European* (pp. 371-375). IEEE.

[7] Savitzky, A., & Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. Analytical chemistry, 36(8), 1627-1639.

[8] Ahmadi, A., Destelle, F., Monaghan, D., Moran, K., O'Connor, N. E., Unzueta, L., & Linaza, M. T. (2015, November). Human gait monitoring using body-worn inertial sensors and kinematic modelling. In SENSORS, 2015 IEEE (pp. 1-4). IEEE.

[9] Alexiadis, D. S., & Daras, P. (2014). Quaternionic signal processing techniques for automatic evaluation of dance performances from MoCap data. Multimedia, IEEE Transactions on, 16(5), 1391-1406.

[10] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. Image Processing, IEEE Transactions on, 13(4), 600-612.

Acknowledgments

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement **FP7-601170 RePlay**.

Figure captions

Fig. 16.1: Layout of the platform for the PLAY&LEARN scenario. **(Vicomtech)**
Fig. 16.2: Placement of the Microsoft Kinect in case of reflective floors. **(CERTH)**
Fig. 16.3: Placement of the Microsoft Kinect for correct use (top view). **(CERTH)**
Fig. 16.4: Combination of modules and components for the coach application of the PLAY&LEARN scenario. **(MIRALab)**
Fig. 16.5: Selection of the preferences in the PLAY&LEARN scenario. **(Vicomtech)**
Fig. 16.6: Global flow of the PLAY&LEARN application. **(MIRALab)**
Fig. 16.7: Several screenshots of the capturing platform for the PLAY&LEARN scenario **(MIRALab)**
Fig. 16.8: 3D avatar pipeline **(MIRALab)**
Fig. 16.9: Overview of the image-based 3D scanner **(MIRALab)**
Fig. 16.10: 3D reconstruction pipeline using image-based method **(MIRALab)**
Fig. 16.11: Example of 3D reconstructed avatar **(MIRALab)**