# An Interaction Mechanism for Virtual Reality Based on Upper Limbs Motions Tracking Using Depth Cameras and Inertial Sensors

Eider Silva * Esteban Clua † Anselmo Montenegro ‡ Daniela Trevisan § Luis Valente ¶

Universidade Federal Fluminense, Institute of Computing, Brazil

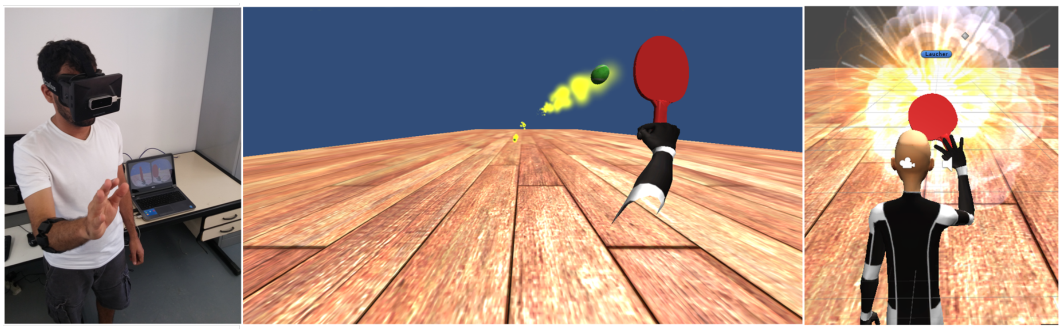


Figure 1: A user playing the VRacket (a), the game virtual environment (b) and the visual feedback that occurs when the user hits a ball (c).

## Abstract

The evolution of the natural interaction between man and computer has represented a positive and promising impact for Virtual Reality (VR) applications. There has been a growing interest in developing new approaches and technologies to improve the user experience so that it can be as natural and immersive as possible. In this context, this work aims to introduce a new concept of natural interaction using the upper limbs with the combination of two types of sensors, classified here as Wearable Inertial Measurement Units (WIMUs) and Head-Mounted Depth Cameras (HMDCs). While HMDCs allow precise tracking of the forearm, hand and fingers, their limited field of view restricts the range of the movements. On the other hand, the WIMUs offer more freedom of movement, since they are not based on cameras and computer vision. However, they are not accurate enough to capture the limbs positions and in details the hands motions. Our solution presents a strategy to combine both classes of sensors in order to improve the user experience with a robust natural interface control. To test the solution, a VR game based on the use of the proposed strategy was developed. An study with tests and evaluation was also developed with users and the results show that the proposed solution outperforms the use of the sensors separately, mainly in terms of performance and fun. Although our proposal is focused on VR games, it can also be an important interaction interface for any other VR based application.

**Keywords:** Virtual Reality, Depth Cameras, Inertial Sensors, Interpolation, Natural User Interaction.

*eidercarlos@gmail.com
†esteban@ic.uff.br
‡anselmo.montenegro@gmail.com
§daniela@ic.uff.br
¶lvalente@ic.uff.br

## 1 Introduction

The upper limbs motion tracking is a very attractive research topic, mainly due to its use in a wide range of applications that are included in a diversity of areas, being explored and studied in the scientific community and applied in the industry. In the search for improvements of related applications, a considerable amount of recent work also combines the use of different types of sensors, taking advantage of the qualities that each one offers and always taking into account the balance between the performance and the cost of acquisition of these devices, in order to always keep the application viable.

With the popularity of VR, the motion capture became more popular which led to the development of commercially affordable devices. However, even the most sophisticated motion tracking device have its pros and cons and their use in virtual environments (VE) still can't provide a natural interaction with a very high fidelity and immersion level.

In the literature, it is possible to find many works dedicated to the use of inertial sensors or depth cameras as the primary way of interaction in a VE. In this work, we classify such sensors according to their features into two main categories: Wearable Inertial Measurement Units (WIMUs) and Head-mounted Depth Cameras (HMDCs). The WIMUs are sensors usually composed by accelerometers, gyroscopes and magnetometers which can be fixed on the user's body to capture motion data from a specific limb [16][1][17][5]. The HMDCs are sensors fixed on a headset in order to track properly the upper limbs, composed in most cases by an infrared projector, an infrared camera, and some are also equipped with an RGB camera [16][11][6][20].

To overcome the limitations found in these classes and improve the tracking system, we propose a sensor integration using a WIMU and a HMDC. The integration is performed with a heuristic which we propose to perform an interpolation-based data fusion.

Varshney in [18] describes sensor data fusion as "Acquisition, processing and synergistic combination of information gathered by various knowledge sources and sensors to provide a better understanding of the phenomenon under consideration". According the author, there are many factors that contribute to the enhancement of the system performance when using data fusion to solve sensors-based problems. We highlight below the main issues which are relevant to our work.

- **Improved system reliability and robustness:** There is inherent redundancy in multi-sensor systems. The system performance is considerably improved due to the availability of data from various sensors. This means that, when one or more sensors fail, the system can continue operating;
- **Extended coverage:** Both spatial and temporal coverage can be increased with the use of multiple sensors;
- **Increased confidence:** The use of more than one sensor assures the increase of the system's confidence, mainly because the sensors can confirm each other's inferences. In the literature we can find examples described in [13][21][14][4];

VR applications require low latency response and immersive interface usage. Sensor integration of optical and inertial based systems may be an important trend for the upcoming industry, being fundamental to address the issues listed above.

Our sensor integration proposal is described in more details in section 3. To evaluate properly the integration proposed, we developed a digital game prototype and elaborated a study to infer the performance and user experience. The details about tests and evaluation are described in section 4.

## 2 Related Works

At the moment of writing this paper, the literature on sensor-based natural user interactions (NUIs) for VR entertainment systems is scarce (e.g. [2]). The related works we found mostly target VR systems for physical rehabilitation (e.g. [3][1][7]).

In entertainment systems, Atienza et al. [2] adapted an existing game to be used with a NUI ("Slash the fruit" for iOS, available in the Apple Store). In the original game, players control an avatar equipped with a sword. To control the sword, players use fast gestures (i.e. touching the screen and dragging his/her finger fast) to cut the virtual fruits that appear on the device screen. Atienza et al. [2] created a NUI that enables players to rotate the avatar using a "head gaze" and to control the sword through hand gestures in the air. The NUI implementation uses the HMD gyroscope to control the avatar rotation and the Leap Motion to detect the gestures. Atienza et al. [2] evaluated their solution with 85 participants and learned that the ease of use was reasonable (M = 3 on a scale from 0 to 5) and the gameplay experience was satisfactory (M = 4.2 on a 0 to 5 scale). As we did in our solution, Atienza et al. [2] also combined different sensors to improve the game interaction. However, we prioritized upper limbs and in terms of motions capture their work is still deficient because of the depth camera (Leap Motion) limitations and the fact that the device must be fixed horizontally in a plane (e.g. a table) - restricting the interaction area.

Baldominos et al. [3] presented a VR soccer game for physical rehabilitation, which uses an Intel RealSense sensor for motion capture. In this game, players take the role of a goalkeeper and uses adduction and abduction gestures to defend balls. There are two ways to conduct rehab sessions, which can be supervised or automatic. In the supervised version, a physiotherapist decides when the ball is thrown and its height in relation to the ground. A faster frequency of launches will require faster movements to be performed by the patient, and a greater difference in height between consecutive launches will require a wider angle in the adduction or abduction movements. On the other hand, the automatic session starts with a longer interval between the launches with a slight difference between the angles, and increases or decreases the difficulty level based on the score. In this way, the game analyzes the user's posture and infers the position of the upper trunk joints (the shoulder, elbow and hand). Therefore, the score is increased only if the movement is performed correctly, which requires that the trunk is straight and perpendicular to the ground, and the arm fully extended. The Intel RealSense depth camera should be fixed in front of the user so that the field of view can cover your entire body. Thus, we observed that although it allows capturing the upper limbs joints (shoulder, elbow and hand), it still has the common limitations of using a fixed depth camera in one position, as mentioned in the work of Atienza et al [2].

Arsenault et al. [1] developed a body tracking system using 10 WIMU sensors attached to major body joints (upper arm, forearm, thigh, shin, trunk, and pelvis) on the right and left sides. This system tracks these joints to move an avatar in real time. To test this system, Arsenault et al. [1] created a VR game where the player's goal is to shoot a target by stretching and pointing the right arm toward it. When the player hits the target, it explodes and the game creates another target in a random location. After the players hits three targets, the targets begin to move continuously, increasing the game challenge. Compared to our work, Arsenault et al. [1] were able to track more body limbs than our system. On the other hand, the system by Arsenault et al. [1] is not able to capture gestures of hands and fingers.

Holmes et al. [7] implemented a VR system for rehabilitation based on physiotherapy using the Leap Motion, Kinect and Myo Armband devices. Called Target Acquiring Exercise (TAGER), the system uses Leap Motion as the main interaction control and aims to motivate the execution of personalized exercises with tasks that stimulate the user to stretch the arm and touch objects of different sizes that appear randomly at different locations in the scene. Holmes et al. [7] conducted a usability study with 23 participants comparing the execution of the tasks wearing the HMD and without it. The experiment results are satisfactory in terms of fun using Leap Motion and report easier tasks when using the VR headset. However, despite using multiple sensors, only the Leap Motion is used for in-game interaction with no integration mechanism. With the exception of the tactile feedback through the vibration (Myo), the Kinect and Myo are only used for storing the produced data to be used in the future.

## 3 Upper limbs motion tracking: design

The main goal of this work is to provide hands-free gestural interaction in VR applications. To achieve this goal our system uses three main hardware: a HMD device to present the virtual world in first-person view, a HMDC device, and a WIMU device, as Figure 2 (a) and (b) illustrates. The HMDC device, which is placed on the HMD, enables the system to track arms, hands, fingers and their respective positions. However, HMDC devices have a limited field of view. To handle this issue, we use a WIMU device to track limbs that are outside the HMDC field of view. This device is attached to the user's right forearm. Although WIMU devices provide only limb rotation and orientation, they are not affected by field of view issues.

### 3.1 Integration Strategy

Due the HMDC accuracy, this proposal has priority to the data produced by it. Thus, while the user's hand is within the HMDC's field of view, the system uses the depth camera to control the avatar's motions. When the hand is outside the HMDC's field of view, the system begins to use the WIMU data. In our proposal, we also dealt with the transition between one sensor and another in order to allow smoothly interpolate the captured data.

The major challenge in implementing this strategy is in the procedures required to deal properly with the transition zone (Figure 2(b)) - whenever the HMDC partially tracks the hand, as when the upper limb starts to fall outside the HMDC field of view. In this

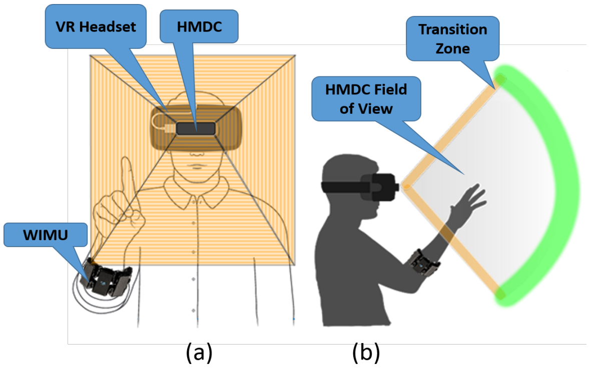


Figure 2: Proposed integration setup viewed from the front (a) and profile (b). The orange area in figure (a) shows the HMDC field of view. The green edge in figure (b) shows the transition zone, where it is performed the interpolation.

situation there are partial data from the HMDC that generate gaps or sudden transitions on the hand poses between the intervals that the device can capture the limbs and the moment that can no longer capture the movements. To cope with this issue, we developed a LERP-based data fusion heuristic (Figure 2). This method smooths the motions at the point of transition through interpolation. In this way, it allows a more natural movement, ensuring a better quality in immersion. After the transition process, the motions are controlled by 100% of the data captured by WIMU or HMDC.

As Kremer [9] describes, the easiest way to interpolate between two positions (points) is using the LERP function. This function has a geometric formula (Eq. 1): given the start ($p_0$) and end ($p_1$) points, and the interpolation parameter $t \in [0,1]$, $Lerp(p_0, p_1, t)$ produces for each $t$ a point along the straight line connecting them. For example, when $t = 0.5$, the result will be a point exactly in the middle of the line between $p_0$ and $p_1$.

$$LERP(p_0, p_1, t) = (1-t)p_0 + tp_1 \tag{1}$$

The proposed integration method is based on the use of the LERP function described above and on the data produced by the WIMU and HMDC devices. The Figure 3 shows in more detail the operation of the proposed method.

In our strategy due to the ability of the depth cameras to produce position data, the upper limb movements of the avatar are based mainly on the IK, where the avatar's wrist is the end-effector and its position is responsible for triggering the arm motions (including the forearm) [8][15].

As seen in Figure 3, when starting the process (application), the procedures are executed within the $LateUpdate()$ function - which implies a function that is executed for each frame. Since WIMU does not rely on any conditions for tracking (just be turned on), for each frame are produced quaternions ($q_0$) - used to calculate the wrist positions ($p_0$) and to move the forearm when the data is being produced by the HMDC. The wrist position ($p_0$) is calculated based on the avatar's forearm size ($s_0$) and direction ($d_0$), and it is used in the interpolation function during the transition process.

Regarding the HMDC tracking procedures - when the upper limb is at its field of view, the data tracking is enabled. Otherwise, it is disabled when the device camera can not identify any limb - which occurs when the user removes the hand or moves the head out of the field of view. The data produced control all the avatar's fingers (as this movement is performed by the user), and as soon as it activates the tracking, it is checked if it was made an interpolation between the position produced by WIMU ($p_0$) and the produced by HMDC ($p_1$). The interpolation function $Lerp(p_0, p_1, t)$ produces $p'$, which is used to move the end-effector through the IK, and consequently triggers the entire avatar arm motions.

The smoothing and stabilization of the motions during the interpolation depends on the $t$ value. Therefore, in order to ensure a good performance, we created the procedure presented in Code 1.

```
1  LateUpdate()
2  {
3     if(isHMDCActive)
4     {
5        if(!isLerpActive)
6        {
7           t = 0.0f;
8           startTime = Time.time;
9           isLerpActive = true;
10          activateInverseKinematics();
11       }
12
13       if(t < 1.0)
14       {
15          timeSpent = Time.time() - startTime;
16          t = timeSpent/0.3f;
17          Vector3 p_ = Lerp(p0, p1, t);
18          wrist.position = p_;
19       }
20       else
21       {
22          wrist.position = HMDC_WristPosition;
23       }
24    }
25 }
```

Code 1: C# Code responsible for defining the value of **t** used in the LERP function.

The Code 1 shows the $LateUpdate()$ function. The lines 7-10 is where is prepared the interpolation process and activated the inverse kinematics function. The condition on line 13 ensures that the value of $t$ (initially 0 on line 7) is incremented until it reaches the value 1 and it is used as a parameter in the LERP function for 0.3 seconds (lines 16 and 17), before exit the interpolation function. In this way, the result produces a smooth motion between the points $p_0$ and $p_1$ during 0.3 seconds. After exit the interpolation, the end-effector (wrist) is moved through the wrist position produced by the HMDC (line 22).

After completing the LERP, the end-effector starts using the value of ($p_1$) produced by the HMDC (while the tracking is enabled). When the tracking is disabled, it is checked whether an interpolation was also done, and in addition, it is checked if $p_1$ is $null$. In this case, $p_1$ should store the last position captured by HMDC, allowing the execution of the LERP to smooth the movement starting at $p_1$ and with the final destination at $p_0$ for 0.3 seconds, which is a procedure similar to that presented in Code 1. When it is identified that $p_1$ is $null$, it means that the capture using the HMDC has not been activated yet, so only the avatar forearm is moved using the quaternion data ($q_0$) produced by the WIMU.

## 4 Implementation and Experiments

To evaluate the user experience with our proposed NUI, we developed a VR ping-pong game, called by VRacket. The main goal in VRacket is to hit as many balls as possible using a small racket, which the player controls by moving his/her arms (Figure 4(a)). The game launches balls towards the player (Figure 1(a)) in random positions and a hit is validated when the launched balls collide with the racket(Figure 1(b)). The launched directions are delimited by a white rectangular area (Figure 4(b)).

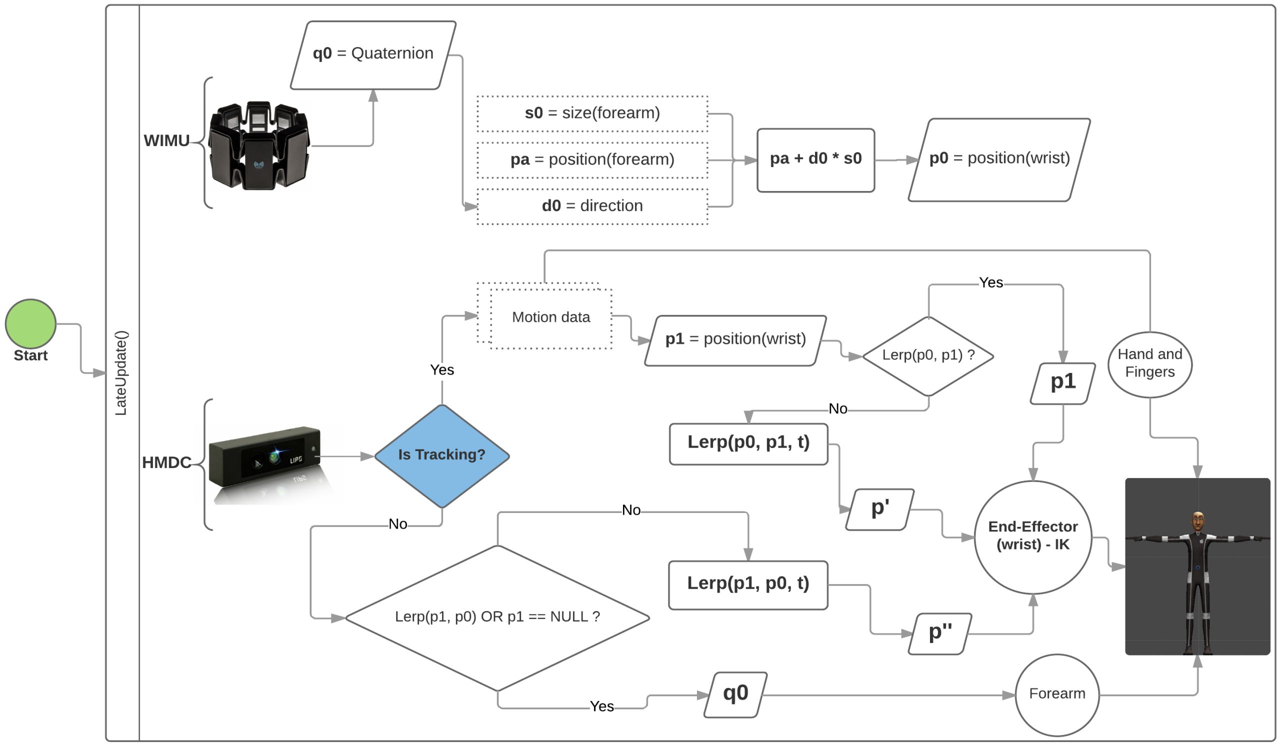


Figure 3: Flowchart of the proposed integration method.

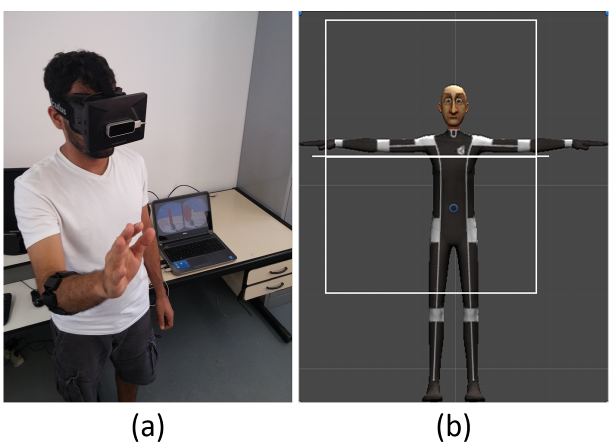


Figure 4: A user performing the evaluation (a) and the rectangle used for delimiting the region where the balls are launched to, which is divided in upper and lower area (b).

The rectangular area in Figure 4(b) is divided into two parts. The upper part represents the area where VRacket is most likey to detect hits using the HMDC (as this sensor is attached to the HMD), and the bottom part represents the area where VRacket is most likely to detect hits with the "WIMU" (as this sensor is attached to the player's right forearm).

We developed VRacket using the Unity3D engine (version 5.5.1f1), with C# to implement the interpolation heuristic. VRacket represents the player avatar as a 3D model with all human skeleton joints, but in our experiments we use only the right upper limb to reproduce the motions capture results.

VRacket runs on a Windows 8.1 64-bit Pro laptop computer, equipped with an Intel Core I7 @1.80 GHz, 8GB of RAM, and a NVidia GeForce GT 740M. The players wear an Oculus Rift DK1 device [19] with an attached Leap Motion [12] (the HMDC device) and a Myo Armband [10] (the WIMU sensor).

### 4.1 Tests and Evaluation

The tests and evaluation intended to compare the performance and gather problematic issues about the user experience using the sensors alone (WIMU and HMDC) and the proposed integration using both. This study was conducted with a total of 20 participants and comprises the following steps: 1) the user must sign a document that informs his/her explicit consent to participate in the study; 2) the user completes a profile questionnaire; 3) we assist him/her in wearing the HMD, having him/her participate in three different test sessions: "WIMU", "HMDC", and "HYBRID"; 4) the user completes a post-test questionnaire about his/her experience.

In the profile questionnaire we asked if the participants had already played games based on movement sensors, such as Kinect (Xbox), Move Controller (PS4), HTC Vive(Steam) or Wii Remote (Nintendo Wii). Twelve users reported having played a few or many times, 5 reported having played once, and 3 reported that they have never played. We also asked if they had already used a VR system. Eight users reported that they had never used this kind of system. Finally, when asked if they had already played ping-pong or tennis, 19 users answered affirmatively.

We designed three different test sessions which are "WIMU", "HMDC", and "HYBRID", and represent the different hardware setups. In session "WIMU", the user uses only the WIMU device (i.e. Myo armband). In session "HMDC", the user uses only the HMDC device (i.e. Leap Motion). Finally, in session "HYBRID" the user uses both devices. The session ordering is random for each user to attenuate possible bias.

Each session starts with a training phase where VRacket launches 15 balls, so that the user is able to get acquainted with the interaction method according to the current test hardware. There is no scoring in training. Next, the test phase begins with VRacket launching 100 balls in random directions inside the area illustrated in Figure 4. The balls are released every 1.5s and each one has speed of 3.5 m/s.

Besides scoring (i.e. amount of hit and missed balls), the test phase records the position of hits and misses to create a heat map (Figure 5), which helps to understand where the users had more difficulty to hit balls according to each kind of test session. This setup enables to compare the pros and cons of the sensor integration we propose in the paper.

After completing the tests the participant answers a user experience questionnaire about ease of use, immersion, performance, discomfort, and fun. These questions are based on a 5-point Likert scale as described below.

**INTERACTION USABILITY EVALUATION**

**[APPROACH]**

1. **How do you rate the ease of use with the device [APPROACH]?**
   Very Easy ( ) ( ) ( ) ( ) ( ) Very Hard
2. **How do you rate the immersion level with the device [APPROACH]?**
   Very Low ( ) ( ) ( ) ( ) ( ) Very High
3. **How do you rate your performance with the device [APPROACH]?**
   Very Low ( ) ( ) ( ) ( ) ( ) Very High
4. **How do you rate the discomfort level with the device [APPROACH]?**
   Very Low ( ) ( ) ( ) ( ) ( ) Very High
5. **How do you rate the fun level with the device [APPROACH]?**
   Very Low ( ) ( ) ( ) ( ) ( ) Very High

These 5 questions were applied in relation to each of the 3 approaches (where was changed the word [APPROACH] to each of the approaches proposed in this work), resulting in a total of 15 questions.

## 5 Results and Discussions

### 5.1 Objective Results

Table 1 lists the average hit rate with the “WIMU” (M=45.70, SD=17.12), “HMDC” (M=45.40, SD=12.63), and “HYBRID” (M=52.45, SD=16.12) approaches. With the “WIMU” approach, the average hit rate in the lower area (M=30.15, SD=11.32) overcame the average hit rate in the upper area (M=15.55, SD=7.45). Concerning the “HMDC” test session, the average hit rate in the upper area (M=31.40, SD=9.23) overcame the average hit rate in the lower area (M=14.00, SD=5.81). With the “HYBRID” approach the results were more balanced: the average hit rate in the upper area was 29.85 (SD=9.56) and in the bottom was 22.60 (SD=10.12). These results suggest that our integration solution worked better than using each sensor alone.

We recorded the positions of all balls (6000) launched in the experiment to generate three heat maps (Figure 5). The heat maps present a visual and spatial feedback about the hit and miss rate. The heat maps also complement the data presented in Table 1.

### 5.2 Subjective Results

In the post-experiment questionnaire, the participants evaluated each test session (“WIMU”, “HMDC”, and “HYBRID”) and answered a final question, which was: “Which of the 3 interaction types did you like the most?”. To evaluate each test session, each participant answered five questions (as a 5-point Likert scale), which were: (1) “What rating you give to the ease of use?”, (2) “How do you rate the immersion level?”, (3) “How do you evaluate your performance?”, (4) “How do you rate the level of discomfort”, and (5) “How do you rate the level of fun?”. The topics in the subsection below summarizes these evaluations.

### 5.3 Discussion

In this section we use “WIMU”, “HMDC”, and “HYBRID” (“WIMU/HMDC”) to refer to the corresponding hardware setups used in the tests.

The quantitative results in Table 1 shows that in “WIMU” the participants hit more balls in the lower region when compared to the upper region (the rectangle in Figure 4(b). In “HMDC”, the hit rate was higher in the upper region when compared to the lower region. On the other hand when using our “HYBRID” approach the participants had a more balanced average hit rate (considering both regions), still being slightly higher in the top region (difference of 7.25). Figure 5 illustrates these situations. In particular, in Figure 5(c) (heat map of “HYBRID”) blue color dots (success rate) are denser at the top when compared to the bottom. We observed that this result was mainly due to the WIMU sensor accumulating errors and losing calibration over time. The participants handled this situation by performing moves within the HMDC sensor field of view. We have tried to use the Kalman Filter algorithm to filter noise and to correct error accumulation, but this approach degrades performance severely (e.g. high frame rate drop), which is unfeasible for our proposal.

#### 5.3.1 Ease of Use.

As shown in Figure 6 concerning the “WIMU” approach, 55% of the users rated this approach as having “medium” ease of use. Considering “HMDC”, 40% of the users considered this approach as “easy”, and 45% of the users considered the “HYBRID” approach as “easy”. When evaluating mean values based on the 5-point Likert scale, users considered “HMDC” as the easiest to use (M=2.80), followed by the “HYBRID” approach (M=2.90), and the “WIMU” variant (M=3.05).

We infer that “HMDC” overcame the other alternatives due to its high accuracy to track hands and to reproduce the avatar moves in real-time with high fidelity, which makes it easier to hit a ball. The “HYBRID” approach had less fidelity due to the bottleneck in the transition area (i.e. the intersection between the HMDC sensor and the WIMU sensor). The interpolation process is performed in the transition moment and during this process the user may perceive some glitches in the avatar animation, which may affect the ease of use negatively. However in “WIMU” the ease of use is worse due to loss of calibration over time, which causes the system to render the avatar forearm in incorrect positions when compared to the user forearm, generating inconsistencies.

Table 1: Objective Results

| APPROACH | REGION | AL | AH (%) | M | SD |
|---|---|---|---|---|---|
| **OVERALL** | **Total** | 6000 | 2871 (47.85%) | 47.85 | 15.75 |
| | **Upper** | 3012 | 1536 (51%) | 25.60 | 11.32 |
| | **Lower** | 2988 | 1335 (44.68%) | 22.25 | 11.47 |
| **WIMU** | **Total** | 2000 | 914 (45.7%) | 45.70 | 17.12 |
| | **Upper** | 1049 | 311 (29.65%) | 15.55 | 7.45 |
| | **Lower** | 951 | 603 (63.41%) | 30.15 | 11.32 |
| **HMDC** | **Total** | 2000 | 908 (45.4%) | 45.40 | 12.63 |
| | **Upper** | 980 | 628 (64.08%) | 31.40 | 9.23 |
| | **Lower** | 1020 | 280 (27.45%) | 14.00 | 5.81 |
| **HYBRID** | **Total** | 2000 | 1049 (52.45%) | 52.45 | 16.12 |
| | **Upper** | 983 | 597 (60.73%) | 29.85 | 9.56 |
| | **Lower** | 1017 | 452 (44.44%) | 22.60 | 10.12 |

AL = Amount Launched, AH = Amount Hit, M = Mean and SD = Standard Deviation.

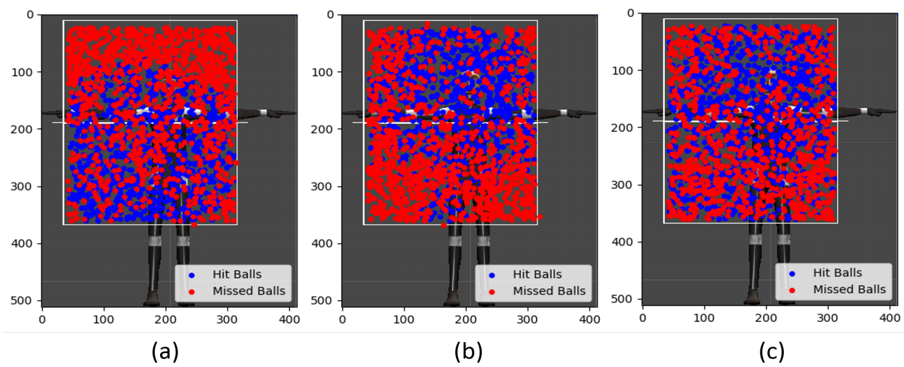


Figure 5: Heat maps with the hit and miss rate using WIMU sensor (a), HMDC (b) and the HYBRID proposed integration (c).

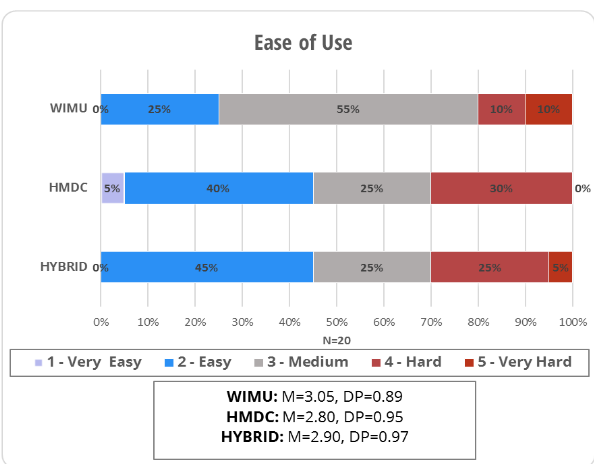


Figure 6: User experience evaluation results in relation to the ease of use.

### 5.3.2 Immersion Level.

According with the Figure 7, a total of 55% of users rated the immersion level in "WIMU" as "medium". On the other hand, users rated immersion level as "high" in "HMDC" and "HYBRID" approaches with similar values (35% and 40% respectively). As 20% of users rated immersion level in "HMDC" as "very high", "HMDC" overcomes our "HYBRID" approach in this item by a slight difference of 0.05 on average.

The immersion level is better in "HMDC" due to the same reasons we described for the ease of use. Although this experiment does not use the user fingers in interactions, the depth camera is able to track finger joints more accurately than using only a inertial sensor (WIMU). This feature of HMDCs enable the virtual environment simulation to be richer in terms of fidelity.

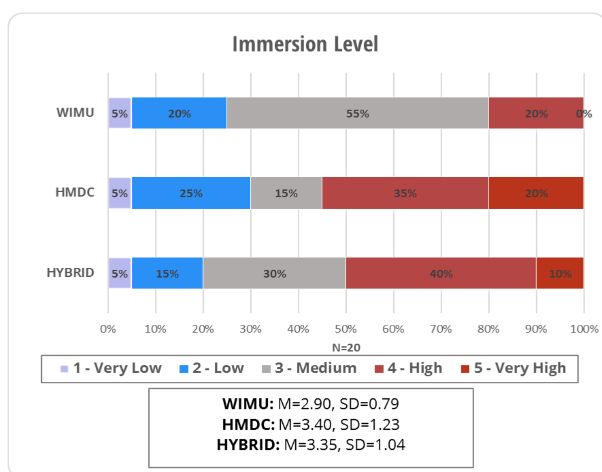


Figure 7: User experience evaluation results in relation to the immersion level.

#### 5.3.3 Performance Level.

Concerning "WIMU" and "HMDC", the Figure 8 shows that just 15% and 10% of the users (respectively) rated these approaches as "high" in this item; no user rated it as "very high". On the other hand a total of 30% of users rated our "HYBRID" approach as "high" and 15% as "very high" in this item, which was decisive to consider that our method outperforms "WIMU" and "HMDC". Our "HYBRID" approach outperforms the others both in the quantitative results and in the user experience questionnaire. These results reinforce the robustness of the "HYBRID" approach.

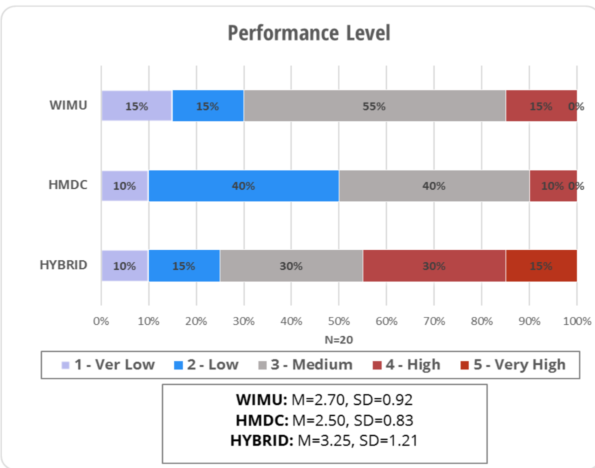


Figure 8: User experience evaluation results in relation to the performance level.

#### 5.3.4 Discomfort Level.

As shown in Figure 9, the most uncomfortable setup was "HMDC" (M=2.15), followed by our "HYBRID" approach (M=2.10). Surprisingly, users rated "WIMU" as the most comfortable to use (M=1.85).

We concluded that "HMDC" being the most uncomfortable stems from requiring the user to position his/her hand continuously in front of the "HMDC" field of view, which generates physical fatigue. We also observed that room illumination interferes with the "HMDC" tracking process, due its infra-red based approach. In particular, high illumination levels generate noise in the tracking process. Users handled this issue by repositioning his/her arms continuously until the device was able to track the limbs properly. This issue also happens in the "HYBRID" approach. The "WIMU" approach simply reproduced the user limb orientation in the virtual environment, despite noise and accumulation errors. It was an unexpected result due the fact that the WIMU sensor must to be fixed and can tighten the user's forearm causing some discomfort.

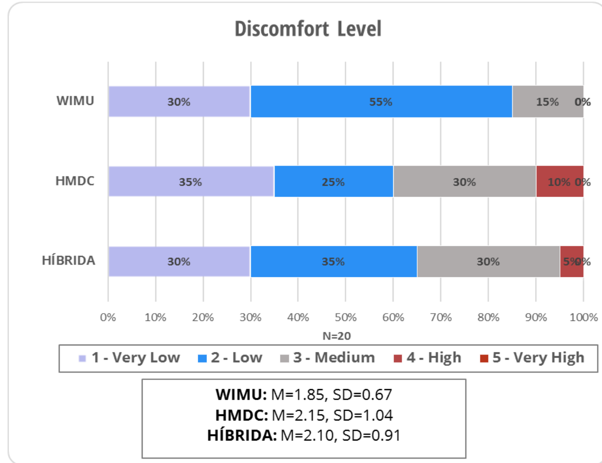


Figure 9: User experience evaluation results in relation to the discomfort level.

#### 5.3.5 Fun Level.

Considering our "HYBRID" approach, the Figure 10 shows that 30% of users rated it as "high" and 25% as "very high" in this item. This latter score affected the overall results on fun level, which we concluded as users enjoying this approach more than "WIMU" and "HMDC". In "HMDC", 30% of users rated it as "high" and 20% as "very high" regarding fun, and in "WIMU" 35% of users rated it as "high" and 10% as "very high".

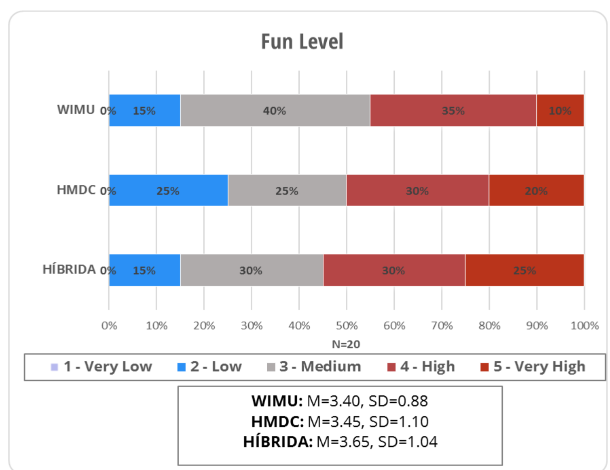


Figure 10: User experience evaluation results in relation to the fun level.

#### 5.3.6 Most liked interaction.

According to our users, our "HYBRID" approach overcame others in preference (Figure 11), with 70%. The "WIMU" variant outperformed "HMDC" by a difference of 10%.

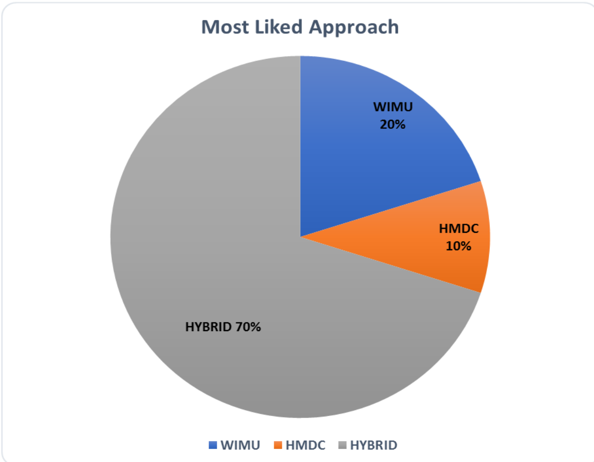


Figure 11: User experience evaluation results in relation to most liked approach.

## 6 Conclusion and Future Work

In this paper we present the design and evaluation of a proposed solution with the goal of combine the use of different sensors (inertial and visual) to capture movements of the upper limbs, allowing a more robust interaction in VR environments.

An integration method based on the linear interpolation function (LERP) was implemented, aiming to smooth the motions in the transition zone, which is at the limit of the HMDC field of view.

Based on the integration method, an experimental game for VR, called VRacket, was developed. The game is based on tennis and ping pong, and requires users to move their upper limbs quickly to hit the highest amount of balls thrown in random regions. This system can also be easily adapted for upper limb physical rehabilitation applications.

The developed game was submitted to evaluation of efficiency and usability with users, in order to validate the proposed interaction and to evaluate comparing with the alternative proposals - which implies the use of the devices (WIMU and HMDC) separately. In this evaluation, an important factor is the difference in the performance of the users when hit the virtual balls launched in the upper and lower body areas (Figure 4(b)), in relation to the approach used during the experiment.

According to experiment results and user experience evaluation, our proposed solution ("HYBRID" approach) outperformed other alternatives that use only one type of device (HMDC and WIMU). Therefore the results are summarized by the following statements:

- The hit rate at the top is higher in "HMDC";
- The hit rate at the bottom is higher in "WIMU";
- The hit rate at the top is much higher in "HYBRID" when compared to "WIMU";
- The hit rate at the bottom is much higher in "HYBRID" when compared to "HMDC".

For future works we plan to improve the algorithm to filter IMU sensor data in order to cope with the noise and errors accumulation, which affects the simulation negatively. We also plan to improve the algorithm responsible for the sensor data fusion between inertial sensors and depth cameras, which is a current bottleneck in our solution.

## References

[1] D. L. Arsenault and A. D. Whitehead. Quaternion based gesture recognition using worn inertial sensors in a motion tracking system. In *Games Media Entertainment (GEM), 2014 IEEE*, pages 1–7. IEEE, 2014.

[2] R. Atienza, R. Blonna, M. I. Saludares, J. Casimiro, and V. Fuentes. Interaction techniques using head gaze for virtual reality. In *Region 10 Symposium (TENSYMP), 2016 IEEE*, pages 110–114. IEEE, 2016.

[3] A. Baldominos, Y. Saez, and C. G. del Pozo. An approach to physical rehabilitation using state-of-the-art virtual reality and motion tracking technologies. *Procedia Computer Science*, 64:10–16, 2015.

[4] S. Ceriani, C. Sánchez, P. Taddei, E. Wolfart, and V. Sequeira. Pose interpolation slam for large maps using moving 3d sensors. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 750–757. IEEE, 2015.

[5] P.-J. Chen, Y.-C. Du, C.-B. Shih, L.-C. Yang, H.-T. Lin, and S.-C. Fan. Development of an upper limb rehabilitation system using inertial movement units and kinect device. In *Advanced Materials for Science and Engineering (ICAMSE), International Conference on*, pages 275–278. IEEE, 2016.

[6] T. Helten, M. Muller, H.-P. Seidel, and C. Theobalt. Real-time body tracking with one depth camera and inertial sensors. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1105–1112, 2013.

[7] D. Holmes, D. Charles, P. Morrow, S. McClean, and S. McDonough. Usability and performance of leap motion and oculus rift for upper arm virtual reality stroke rehabilitation. In *Proceedings of the 11th International Conference on Disability, Virtual Reality & Associated Technologies*. Central Archive at the University of Reading, 2016.

[8] R. N. Jazar. *Theory of applied robotics: kinematics, dynamics, and control*. Springer Science & Business Media, 2010.

[9] V. E. Kremer. Quaternions and slerp. In *University of Saarbrucken, Department for Computer Science Seminar Character Animation. Retrieved from: http://embots.dfki. de/doc/seminar_ca/Kremer_Quaternions. pdf*, 2008.

[10] T. Labs. Myo gesture control armband — wearable technology by thalmic labs, mar 2017.

[11] A. McWilliams. How a Depth Sensor Works - in 5 minutes, 2013.

[12] L. Motion. Motion controller for games, design, virtual reality and more, mar 2017.

[13] B. Penelle and O. Debeir. Multi-sensor data fusion for hand tracking using kinect and leap motion. In *Proceedings of the 2014 Virtual Reality International Conference*, page 22. ACM, 2014.

[14] A. M. Sabatini. Quaternion-based extended kalman filter for determining orientation by inertial and magnetic sensing. *IEEE Transactions on Biomedical Engineering*, 53(7):1346–1356, 2006.

[15] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1):116–124, 2013.

[16] E. C. Silva, E. W. Clua, and A. A. Montenegro. Sensor data fusion for full arm tracking using myo armband and leap motion. In *Computer Games and Digital Entertainment (SBGames), 2015 14th Brazilian Symposium on*, pages 128–134. IEEE, 2015.

[17] Y. Tian, X. Meng, D. Tao, D. Liu, and C. Feng. Upper limb motion tracking with the integration of imu and kinect. *Neurocomputing*, 159:207–218, 2015.

[18] P. K. Varshney. Multisensor data fusion. *Electronics & Communication Engineering Journal*, 9(6):245–253, 1997.

[19] O. VR. Oculus rift, mar 2017.

[20] X. Wei, P. Zhang, and J. Chai. Accurate realtime full-body motion capture using a single depth camera. *ACM Transactions on Graphics (TOG)*, 31(6):188, 2012.

[21] H. Zhao and Z. Wang. Motion measurement using inertial sensors, ultrasonic sensors, and magnetometers with extended kalman filter for data fusion. *IEEE Sensors Journal*, 12(5):943–953, 2012.