An Augmented Reality U-Academy Module: From Basic Principles to Connected Subjects

Paulo Menezes
University of Coimbra, Coimbra, Portugal
`paulo@isr.uc.pt`

Abstract—A module for learning about virtual and augmented reality is being developed under the U-Academy project. The module is composed of three parts. The first part is an introduction to the basic concepts of virtual and augmented reality with the help of illustrative examples. The second part presents some of the current uses of augmented reality and its prospective use in several areas that range from industry to medicine. The final part aims at those students interested in the insights of this technology by presenting the underlying concepts such as: camera models, computer graphics, pattern detection and pose estimation from inertial sensors or camera images.

1 Introduction

Augmented reality has recently received an enormous amount of attention from both the general public and companies. Naturally, the game industry has been quite attentive to the long promised technologies to support it. There are indeed several companies doing important investments in the development of products for supporting augmented reality (AR), like Microsoft Hololens, Vufuria, Magic Leap, and Meta 2, or for virtual reality (VR) such as HTC Vibe and Oculus Rift.

As a matter of fact, some technical difficulties have limited the achievable quality of visualization in AR, and for this reason its inclusion in the games offered by the major players in this industry has been postponed until recently. But, with the new visualization devices available, and the high computational power of current game consoles and personal computers, we can say that the principal barriers to AR adoption have been removed. Moreover, the current awareness of the general public to this type of technology makes it impossible to ignore at the risk of losing visibility to other competing companies.

While on the users side it is mainly the novelty that attracts attention, in particular amongst the younger generations, on the commercial side several companies have perceived the possibilities that this new concept was creating. As a result, we have seen the promotion of products via AR, for example by adding markers to their en-

closing boxes that can be used with some AR-enabled applications, typically downloadable to smartphones and tablets.

Beyond the promotional use, there are indeed several areas where augmented reality may create new opportunities and added value. Fashion selling stores can use it for enabling people to try on clothes without having to put them on and off. We can also expect that it will bring important benefits to several industrial areas, and in particular to manufacturers, that have the opportunity to include it as a helping tool in assembly, inspection, or maintenance tasks. Among the foreseen uses, we can mention the use of AR for providing guidance about the sequencing of operations to be executed during the inspection of aircrafts, complex assembly procedures, or maintenance tasks. Beyond guidance, it may support the visualization of quantities being measured at a given instant, or related with some functioning parameters of a particular machine.

For all these reasons, it becomes clear that engineering students should be introduced to the AR concepts, as it is most likely that they will encounter this type of technology in their future workplace. Beyond the question of what AR is, how it differs from VR [1,2,3], and how it can be used, the question of what is it built upon may also be explored, either by the curious student, or in the context of specific courses like computer vision (CV) or human-machine interaction (HMI).

The use of AR as a motivation for computer vision can be employed to give practical examples of the use of various subjects that may range from pattern recognition to projective geometry. In the case of HMI, it opens the possibility to use AR as a basis for the creation of new interaction mechanisms. These new mechanisms in turn may be applied to support activities like: AR-guided minimally invasive surgery, immersive teleoperation of micro or remote robots, tele-surgery and tele-diagnosis, to name but a few.

The remainder of this paper will present some of the subjects that will be progressively integrated on this U-Academy module. The next section discusses some concepts that are needed to understand the difference between AR and showing information or graphics on top of images, how is reality perceived and what are the ingredients for creating systems capable of inducing augmented reality perceptions. Section III provides an analysis of the main types of interaction used nowadays, their limitations and the need to develop direct manipulation mechanisms. Section IV is about the development of AR applications and how it can be explored for motivating students to subjects like computer vision, signal processing, filtering and estimation, graphics programming, or even electronics. Related with the two latter subjects examples are presented of inertial-based hands, and object trackers that can be used to explain both the electronics, the signal filtering and estimation processes or even computer graphics. Section V summarizes and concludes the article, by leaving pointers for the interested reader to access the material of the module that is already available.

2 Augmented Reality Concepts

There are indeed several misconceptions about augmented reality (AR), especially among programmers and companies willing to use the current hype to promote their products. The most common one is the notion that for creating AR, one needs to get some nice 3D model and just superimpose it on live video. In fact, that can be part of it, but it is not enough to create "augmented reality. This is similar to the subtitles that frequently appear superimposed on a movie or TV show, but which are not (perceived as) part of the scene (or "reality") being shown.

On the other hand, it has become common in sport transmissions to have virtual field marks displayed on the field, e.g. to help spectators understand why a referee has taken some decision, or why someone claims that it was a wrong decision. In these cases, those marks can be perceived as lying on the field, so they "augment" the perceived scenario. For this reason, we can say that this case corresponds to an example of augmented reality.

2.1 So, what is reality augmentation?

To know how to augment reality we need first to understand what is reality. Is it some absolute truth or is it the result of a set of cognitive processes that involve learned concepts, mental models and perception mechanisms?

As human beings, we can only verify (and accept as true) what we see, touch, hear, smell, or taste, and compare it with memories of previous experiences or with acquired concepts. We can say that it is the combination of what is acquired through the senses, its processing, and matching against pre-learned models that results in the perception of reality. In fact, it also involves the use of pre-acquired models and concepts, that may completely change the interpretation of any sensed (acquired) information.

An example of how knowledge may affect reality perception can be when an adult and a child walk on a field and encounter a strawberry poison-dart frog (Oophaga pumilio). The child will probably become excited with the beauty of the frog and will want to try to catch it, while the adult will be terrified and will stop the child from doing the probably mortal move. Here the two persons will have completely different notions of reality for exactly the same situation.

2.2 Cognition and perception of the reality

Our senses and cognitive processes being limited both in acquisition and processing terms, we have developed impressive capabilities of inference, recognition and reasoning, even in face of incomplete data. This is probably the result of our evolution as to what concerns anticipating dangers or survival advantages. The capability of using partial data has made possible the development of our visual system, which is based on 2D projections of the 3D world, and, from these 2D representations, is able to infer about the 3D structures and deal with them. But the 2D nature of this percep-

tual system leads to the appearance of illusions, that are just the result of some model fitting process upon incomplete or ambiguous data.

Although the two-eye configuration has an important role in the perception of 3D structures, the great capability of our brain for integrating sensory information along time enables us to use self-motion to get more information about the neighbouring 3D structures, in particular when the stereo-based vision is not enough for that purpose. These movements, which are frequently done in an automatic and unconscious way, have the purpose of removing ambiguities or breaking up misinterpretations. In other words, this is the way we check how realistic is what we perceive. This can be seen as a geometry-related consistency verification, where we move to check if the 2D structure we are perceiving respects some 3D mental model that was selected as hypothesis.

2.3 Augmenting (the perceived) reality

To produce augmented reality, it is necessary to generate the required sensory stimuli, through the use of some mediating technology, for enabling the perception of virtual elements perfectly integrated with the real (physical) ones. Being our perception able to extract geometric relationships, it is fundamental that the integration of the virtual models and the "real" scene exhibit spatial coherence. Thus, for an augmented scene to be credible (or realistic), the virtual elements must always appear in the same relative positions and poses with respect to the physical ones. Or, as an observer moves towards, away or around the scene, the view of all virtual and physical elements must suffer exactly the same perspective and rigid transformations. This consistency check enables us to perceive the virtual elements as being part of the scene, and therefore in our vicinity, and as a result we may develop the feeling that we can touch them.

When we achieve a sense of tangibility or sense of presence, as defined by Sheridan [4], we can say that we tend to accept the scenario as real, but for that to happen it must pass all the voluntary and involuntary consistency checks we perform.

2.4 Head mounted displays versus handheld visualisation

Although it can be discussed if the right way to produce AR is by using HMDs or, in alternative, handheld devices (e.g. tablets, smartphones, or other), both of them have advantages and disadvantages as will be seen.

In fact, an HMD with one or a pair of coupled cameras, that transform it into a see-through device, seems to be the right choice for creating AR experiences. It can enable the user to look in any direction and see the augmented scenario. But a handheld device can also be considered as an instrument that enables us to see through it and obtain different and augmented views of the surrounding environment, similarly to the use of a portable magnifier.

There is no distinction between them in terms of the involved principles. In both cases the device enables the exploration of the surrounding environment and sees it with added contents. The technical needs and difficulties are also similar, both requir-

ing the estimation of the viewing pose with respect to the environment in a perfectly stable way. In addition, the extraction of the 3D structure of the environment may enable the correct management of the occluding interactions between real and virtual elements, but this is still a hard task given the computational difficulties it imposes. As a result, both cases can work well in simplified utilization scenarios like planar surfaces containing detectable markers, or in complex ones for which a priori models of the environment exist and precise localization technologies are in use.

The differences between the two systems are on the application scenarios and therefore not on the involved processing or algorithms. We can say that AR on HMDs is adequate for tasks that require the use of both hands and/or require the visualization from a user-centred perspective. The use of handheld devices can be more favourable for use during short periods of time, so that the AR tool can be picked up, used to examine an object or scene for a few minutes, and then released.

One should note that although AR can make use of different kinds of visual markers to detect and select the information to display, if the visualized object does not appear perfectly integrated in the environment, we cannot say that AR is being used. In such a case, it is just a QR-code (or other) reader application that displays the related information, eventually after fetching it from some database.

We can say that in many cases we do not really need AR, or, even worse, using it may render the task more difficult than operating a simple code reader that selects the appropriate information to show. And the reason is that, in most situations, it is more practical to scan the code and then look at the device display in the normal handling position, than to keep it up in front of some marker for reading the same information. Conversely, AR may be very practical and useful in situations where the device can be used like a hand magnifier and interactively visualize information about objects, devices, or places just by passing the handheld in front of them.

3 Interaction Issues of Augmented Reality and Connected Devices

In most of the current AR applications the interaction is limited to the motion of the handheld device or HMD as a way to change the point of view with respect to the scene that contains the virtual elements. We can say that for several cases this is sufficient if the objective is only the visualization of those elements. But what happens when the user wants to select different types of information, or eventually interact with the virtual elements to modify their behaviour, or even to use them as control inputs for some physical system? The handheld approaches can make use of the touchable interface to select, open menus and select options of these elements. Conversely, the HMD-based applications are typically hand free approaches, where the interactions may be made using buttons on the helmet itself, if they exist, using gamepads or other handled device, by performing some specific "air-gestures", or using any other nearby interaction device. In the HMD-based AR cases there are indeed many possibilities for the interaction, as the surrounding environment remains visible, and so keyboards, button pads, or any traditional device available may be

used. Unfortunately, most of these devices only provide indirect ways of performing interaction, and this is quite far from natural if we thing about opening a (virtual) drawer using a keyboard or a gamepad. And, if instead of one, several virtual objects do exist in view, will we need to memorize all the corresponding buttons? What if each of the *n* objects has *m* degrees of freedom, should the number of buttons correspond to *m×n*, or will we use a selection-before-action method? There is no single solution for this problem, but making the number of controls explode is normally not acceptable as it will oblige the user to learn the mapping and recall it during every interaction session. For this reason, direct manipulation is more favourable as it does not require particular training phases, because acting on the virtual objects is done resembling the usual manipulation of real objects, or through some physical objects that interact with the virtual ones, as is the case shown in Fig. 1.
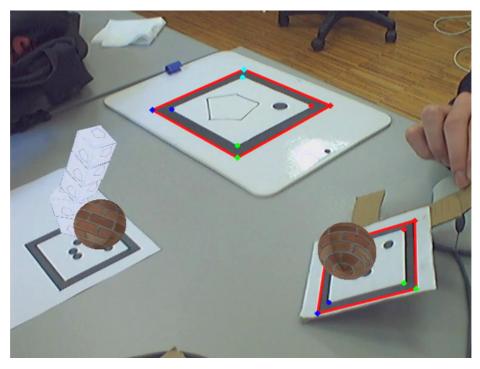

Fig. 1. Interaction example in AR scenario: Varying the inclination of the planar marker to make the ball roll down towards other virtual objects.

3.1 Direct manipulation and sense of touch

All the above-mentioned interaction mechanisms can be used to modify the behaviour of the application, or even that of the virtual objects added to the scene. But how strange it may be to change object properties like the position or the orientation through one of these indirect interaction mechanisms. Our intuition (or mental model) tells us that moving an object can be done by touching, grabbing, moving and then

releasing it, or, in a simpler way, by pulling and pushing it. We can say that the natural and intuitive way of moving (or interacting with) an object is through direct manipulation.

Nevertheless, touching virtual objects is still restricted to the use of rod-like interfaces of haptic devices that enable to touch the objects indirectly with the rod or pencil handled by the user. There are however other works on the development of approaches to produce touch-like sensations, or to modify the touch sensations. Some of the most interesting are the use of vibro-tactile [5] gloves, air flow modulation [6] for enabling the user to perceive, up to some level, the sensation of touching virtual objects, and electrostatic vibration for modifying the perceived texture of a surface [7]. All of these cases need to get a precise estimation of the hand motion, and more precisely of the finger tips, to control the generated stimulus depending on their position.

3.2   Tracking hands

The interaction with virtual objects using direct manipulation has been at the centre of attention of several researchers [8], for the reasons explained above. The main difficulty is how to reliably track hands and their gestures, given the high number of degrees of freedom of their articulated nature, and deal with the frequent occurrence of self-occlusions. Vision- or image-and-depth-based approaches have shown good results, as is the case of LeapMotion, Structure Sensor with OpenNI SDK, or Intel PrimeSense, but they are still limited to configurations where the gestures occur in limited volumes without obstructions. Some attempts to use LeapMotion devices mounted on the HMD have shown some good results, but hands are better tracked from below given that their natural poses generate many occlusions when observed from above.

There are also solutions based on wearable hand trackers, that typically provide very good results. The negative aspects are the inadequacy of the use of gloves for some activities, and the price of these devices.

4   Creating Augmented Reality Applications

To create an augmented reality application, independently of the target device being a handheld or a HMD, the principle is to create the illusion that virtual objects or entities are part of the environment that can be viewed by the user. Excluding glass-like devices because these raise a new set of problems that are out of the scope of this paper, the remaining systems employ one or two cameras to capture the view of the environment, and that view is going to be exhibited on the device display. To create the reality augmentation effect, new virtual elements are introduced in the viewed scenario, through their combination with the captured images. This mixing of virtual and real elements should be sufficient to create the intended perception. But for this to be true, the appearances of both virtual and real elements must evolve along time in exactly the same way, i.e. suffer the same viewing transformation.

For the purpose of generating those viewing transformations, the relative pose of the real objects with respect to the viewing camera, must be known. This relative pose is one of the necessary set of parameters, the second set being the intrinsic parameters of the real camera. While the first serves to set the pose of the virtual objects in the virtual camera referential, the second can be used to define a matching projection matrix.

This estimation can be done by using any type of technology that enables to track both parts, or just one with respect to the other. Although several types of technology are available, their high costs make them prohibitive for consumer grade products. There are however two that, being inexpensive, are typically used for this purpose, being based on: 1) detecting and tracking markers for pose estimation using the camera images, and 2) using sensors that provide measures of displacement- and rotation-related quantities, e.g. inertial magnetic units (IMU).

Both have advantages and disadvantages that can be summarized as follows: visual marker-based pose estimation is simple and provides stable pose estimates, but it is affected by illumination variations, and does not typically behave well when markers are not fully visible, either making the virtual models appear or disappear instantly, or stop moving thus not following the camera/marker movements.

On the other hand, IMUs do not provide direct pose information, and it has to be estimated using integration of angular velocity for orientation and double integration of the measured acceleration for position. The problem is that this type of estimation tends to drift, given the accumulation of some minor bias that may be present on the measures. Fortunately, in what concerns orientations, it is possible to obtain quite reliable results through the combination of estimated gravity acceleration, angular velocity and Earth magnetic field orientation. For this reason, IMU-based AR applications typically only use the estimated orientation of the camera to manipulate the view, not allowing for lateral, up-down, or proximity changing movements.

4.1 Direct manipulation for virtual objects

To interact with the virtual elements using direct manipulation, there are several possibilities, but we can say that intuition normally tells us that if we are seeing the objects in front of us, we should be able to reach, move, or act on them with a hand, a finger, or some held object.

The simpler solution is, indeed, the use of some object, e.g. plier, forceps, or stick, for the interaction. This object will be a physical object manipulated by the user, and by being tracked, a virtual representation may be generated and managed accordingly.

In contrast, the direct interaction with the user hands is typically a more complicated solution, given the high number of degrees of freedom, and self-occlusions constantly generated that affect trackers based on hand images.

4.2 A wireless hand tracker

Given the importance of capturing the hand movement, not only for AR but also for VR, several solutions were investigated and as a result a low-cost wireless hand

tracker was developed at the Institute of Systems and Robotics of the University of Coimbra (ISR-UC). This new device comprises a set of 6 hardware modules: one main board that connects the five other smaller boards, one for each finger. Each board contains an inertial measurement unit (IMU) and through the use of a special purpose adaptation of the Complementary Filter [9] estimation algorithm it is possible to track the hand and its finger movements. The constructed prototype is shown in Fig. 2 (left), where the parts are individually identified.
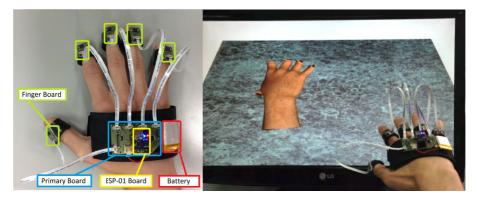

Fig. 2. Left: Prototype of hand motion tracker based on MPU-9150 and ESP-01 wireless processing board; Right: Example of hand tracking and model animation.

As the motion capture provided by this device is based on inertial sensors, only finger flexion-extension (adduction-abduction) and hand orientation are considered, given that, as previously mentioned, position (translation) estimation normally suffers from error accumulation that typically makes it unusable after a few seconds. The principles and description of the design of this device are available on the website. This device, being accessible through a TCP/IP connection, can be included in different types of applications that require the capture of hand motion, eventually for both hands by using two devices.

Being wireless it can be used in a variety of applications, namely in AR or VR. In particular, if used in conjunction with a Kinect sensor, it is possible to capture full body and hand movements, even in configurations where the hands are occluded by other body parts or objects.

Fig. 2 (Right) shows an image where a virtual hand replicates the user hand movements captured by the device. As this device is simple to use and does not create any kind of constraint to the hand movement, it has the right characteristics for interacting with virtual objects in AR, eventually complemented with vibro-tactile stimulation for creating the illusion of touching or holding those objects.

4.3 Touchable but virtually modifiable objects items

One of the features that defy realism, when dealing with the manipulation of virtual objects, is the lack of sense of touch. Despite the attempts in perfecting haptic devic-

es, these still have limitations in terms of the provided stimuli, and manipulation constraints.

To circumvent this limitation, our team is creating physical objects that can be easily tracked or even instrumented. As a result, these objects that may be tracked while handled may therefore become active in the AR scenarios, and be represented by superimposed models that aim at changing their perceived appearance, or simply be used to interact with virtual models. The interaction with virtual models may be very interesting for explaining some physical laws to students, as shown by the previous inclined plane example. In complement, using the appropriate sensors, it is possible to give controllable perceptions of an object in terms of some of its properties like rigidity, as proposed by Restivo et al. [10].
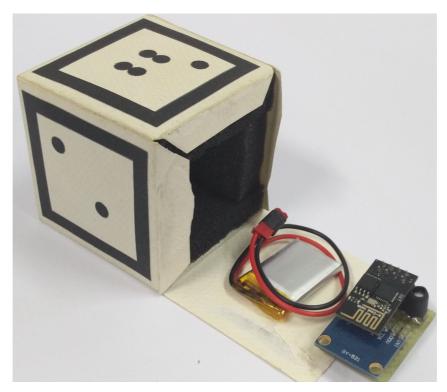

Fig. 3. An instrumented cube, whose faces are covered with visual markers for pose estimation, equipped with a wireless processing board, IMU, and pressure sensors.

To enable its use alone or to improve it in conjunction with a hand tracker to extend its use and enable capturing of not only orientation changes but also displacements, we included a set of easily distinguishable and identifiable markers on the object surface. Using a camera to capture images of this object, these markers can then be detected and used to estimate the object pose using a computer vision approach. Although this pose estimation is only possible when at least one of the markers is visible, the associated errors are due to the discrete nature of the images and do

not accumulate over time, as in the inertial-based case. Since the object also contains accelerometers, gyroscopes and magnetometers, a fusion algorithm can take advantage of both methods. In fact, the visual marker provides a stable estimate of the object pose but may produce spurious erroneous estimates due to numerical problems or motion blurred images, while the IMU provides better information about its movements, in a smooth but increasingly biased estimate. Therefore, a combination of the two estimates can improve the quality of the result and consequently the generated representation of the virtual model.

It should be noted that, as the camera images are normally produced at a lower rate than the inertial measures, the fusion algorithm must take this into account.

4.4 From augmented reality to connected subjects

It is well known that nowadays students are not attracted to technology as they were in the twentieth century. This has led to reflexions, discussions and experiences on how to motivate the young generations to science, technology, engineering and mathematics (STEM). While in the past the normal way of learning started by the basic subjects, currently there are pedagogical experiences on some subjects with interesting results based on top-down approaches, i.e. starting with the visible face of some technology and then going deeper and deeper towards the supporting concepts. One of these subjects is computer network courses, where instead of using the traditional syllabus that started from the electrical signals modulation, to the bit transmission, and climbing up the protocol stack, there are recent ones that start at the user level services like web page access and then go down the protocol layers: transport, network, data link, and physical. This has the advantage of rooting on a common ground that every student has or can be given access to.

The idea of this module is to do the same around Augmented Reality. Benefiting from the attraction that this subject creates on students, we may use it to motivate the study of several other supporting topics. Hereafter follows a non-exhaustive list of these areas, some of them having been addressed along the previous sections:

- Human Factors: perception, attention, memory, recognition, etc.
- Human Machine Interaction: design of interactive systems
- Computer Vision: pattern detection, geometric transformations, homography estimation and decomposition, camera calibration, etc.
- Computer Graphics: 3D models, rendering, stereo displays, shaders, etc.
- Signal Processing: IMU noise reduction, high-pass and low-pass filters
- Estimation: Kalman filters
- Electronics: development of interaction devices.
- Data Communication: TCP/IP programming for communication between devices

Besides the supporting principles and technologies, AR and VR can also be used to create experiences that support the study of several other subjects. Some of these experiences would not be otherwise accessible to students due to the involved costs or risks, or even unavailability of equipment.

5 Conclusion

The concepts presented in this article are available in a U-Academy module about the subject "Augmented Reality: From basic principles to connected subjects" [[1]1]. There are indeed many other subjects that can be studied making use of the connection they have with AR. As shown above, the establishment of connections between them can be used to motivate the students to study the supporting principles, in order to better understand how they can be used to create AR applications, for example. The module is expected to continue evolving, and eventually grow beyond the supporting principles and technologies to the application fields and areas that already are expected to benefit from it. Besides text, it will integrate images, videos, pointers to demonstrators and example code pieces to enable the students to learn about the principles that support the construction of AR-based applications.

6 Acknowledgment

7 References

[1] P. Milgram and A. F. Kishino, Taxonomy of Mixed Reality Visual Displays, IEICE Transactions on Information and Systems, E77-D(12), pp. 1321–1329, 1994

[2] Jun Rekimoto and Katashi Nagao, The World through the Computer: Computer Augmented Interaction with Real World Environments, Proceedings of the 8th Annual ACM Symposium on User interface and Software Technology. UIST '95. ACM, New York, NY, 29-36 https://doi.org/10.1145/215585.215639

[3] Mann, S., & Fung, J. (2001). Videoorbits on EyeTap devices for deliberately diminished reality or altering the visual perception of rigid planar patches of a real world scene. Proceedings of the Second IEEE International Symposium on Mixed Reality, pp 48-55, March 14–15, 2001

[4] Sheridan, Thomas B, "Musings on telepresence and virtual presence," in *Presence: Tele-operators & Virtual Environments*, Number 1, Vol 1,MIT Press, 1992, pp. 120-126.

[5] Giannopoulos, Elias, Ausias Pomes, and Mel Slater. "Touching the void: exploring virtual objects through a vibrotactile glove." *The International Journal of Virtual Reality* 11.2 (2012): 19-24.

[6] Sodhi, Rajinder, Ivan Poupyrev, Matthew Glisson and Ali Israr "AIREAL: interactive tactile experiences in free air." *ACM Transactions on Graphics (TOG)* 32.4 (2013): 134.

[7] Bau, Olivier, Ivan Poupyrev, Ali Israr and Chris Harrison, "TeslaTouch: electrovibration for touch surfaces." *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. ACM, 2010. https://doi.org/10.1145/1866029.1866074

[8] Vasco Rodrigues, Daniel Mendes, Alfredo Ferreira, Joaquim Jorge, "Mid-Air Manipulation of 3D Models in (Semi-)Immersive Virtual Environments," 22 Encontro Português de Computação Gráfica, 12-13 November 2015, Coimbra, Portugal

[9] Euston, Mark, Coote, Paul, Mahony, Robert, Kim, Jonghyuk, Hamel, Tarek, "A Complementary Filter for Attitude Estimation of a Fixed-Wing UAV", IEEE/RSJ International Conference on Intelligent Robots and Systems, 2008

[10] F. Carneiro, B. Santos, P. Abreu, M. T. Restivo, "A tool for grip evaluation and learning," 13th International Conference on Remote Engineering and Virtual Instrumentation (REV), 2016. https://doi.org/10.1109/REV.2016.7444489

[11] Paulo Menezes, Augmented Reality: From basic principles to connected subjects, online, URL: http://orion.isr.uc.pt/pm/index.php/augmented-reality-module

8   Author

Paulo Menezes is with the Department of Electrical Engineering, and the Institute of Systems and Robotics of the University of Coimbra, Polo II, 3030-290 Coimbra, Portugal, and a member of IEEE. His research interests include computer vision, human-robot interaction, human behaviour analysis, and human perception issues and reactions in immersive systems.