The Role of Augmented Reality in Smart Home Settings

Andrea Mattioli [1], Marco Manca [1], Fabio Paternò [1] and Carmen Santoro [1]

[1] CNR-ISTI, HIIS Laboratory, Via G. Moruzzi 1, 56124 Pisa, Italy

Abstract

Augmented Reality (AR) is a growing trend in technology with countless applications in different domains. However, not much attention has been devoted to the smart home setting and how its application can be used to allow users to customise their living spaces. In this paper, we describe the implementation of two methods for recognising objects with AR, and how an End-User Development (EUD) approach to a smart home can take full advantage of these techniques to provide personalisable and more meaningful experiences to users.

Keywords

Smart Home; Augmented Reality; Object Detection; Internet of Things; End-User Development.

1. Introduction

The interest in AR has grown in recent years. The idea of AR is to enhance the environment currently in the user's field of view, creating a space in which components of the digital world are naturally blended in the user's perception of the real world. Such augmentation often occurs via superimposing interactive virtual information on a user's view of the real world [8]. Researchers have focused on applying AR in diverse domains, such as language learning, older adult home environment safety, and localised information (e.g., in [3,7,8]). However, despite having appealing qualities, AR is still not widespread. Also, not much research attention has been devoted to domestic use cases, a domain that seems to garner user appreciation from an explorative study [6]. Until now, most AR applications use visualisations just to provide information about the characteristics of the objects, whereas applications that offer control capabilities, such as the possibility of configuring automation, are still rare. An example in this sense is [9], which allows for process modelling between IoT objects by framing and drawing connections between them. The application of AR to EUD for the smart home can be a valid research direction because it can reduce the distance of the mapping between the physical objects available in the real environment and their digital counterparts [1]. The AR technology allows a more situated and personalisable approach in terms of the representations used to coordinate and automate their concerted behaviour (e.g. trigger-action rules), by using actionable links between the two. Different devices can be used to render the augmentations in AR, such as glasses, head-mounted displays, wearable, and smartphones. Mobile Augmented Reality (MAR) is a promising field of mobile applications [2]. A personal device can enable new experiences for the users without the burden of purchasing and familiarising themselves with a new device. AR applications can use a real-world point of interest or a feature (such as a marker or a detected object) as a starting point for the augmentation. One of the first things to consider when implementing an AR application that uses real objects as anchors for virtual information is how to detect these objects. This paper describes two possible approaches, one based on the Vuforia object recognition and one on a machine learning object detection model. In the comparison section, the specific aspects of the two approaches are considered. Next,

EMPATHY: Empowering People in Dealing with Internet of Things Ecosystems. Workshop co-located with INTERACT 2021, August 30, 2021, Bari, Italy
EMAIL: andrea.mattioli@isti.cnr.it (A. Mattioli); marco.manca@isti.cnr.it (M. Manca); fabio.paterno@isti.cnr.it (F. Paternò); carmen.santoro@isti.cnr.it (C. Santoro)
ORCID: 0000-0001-6766-7916 (A. Mattioli); 0000-0003-1029-9934 (M. Manca); 0000-0001-8355-6909 (F. Paternò) 0000-0002-0556-7538 (C. Santoro)
© 2021 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-3053/paper_4.pdf

different use cases for an application that uses object detection and AR capabilities in a smart home setting are presented. Finally, we draw some conclusions on the experience so far and foresee some future work.

2. Vuforia-based solution

Vuforia[2] is an AR Software Development Kit that makes available to developers a set of technologies to create a digital representation of real objects (Target Images), whose feature points are used to recognise and track digital objects from the camera data, and to deploy visualisations according to the position and orientation of the camera. The representations of real objects can be generated using the Vuforia Object scanner application, a CAD model, or a 3D scanning of the object. Each approach has distinct features: for small objects such as IoT devices, the Vuforia scanner seems the most suitable one if the use of modelling software is not possible. In the documentation[3], guidelines are specified to make sure that the recognition of objects acquired with the Vuforia scanner works well; namely, objects should not have moving parts, and their surfaces should have a recognisable texture or many contrast-rich features. Also, having a peculiar shape improves recognition.

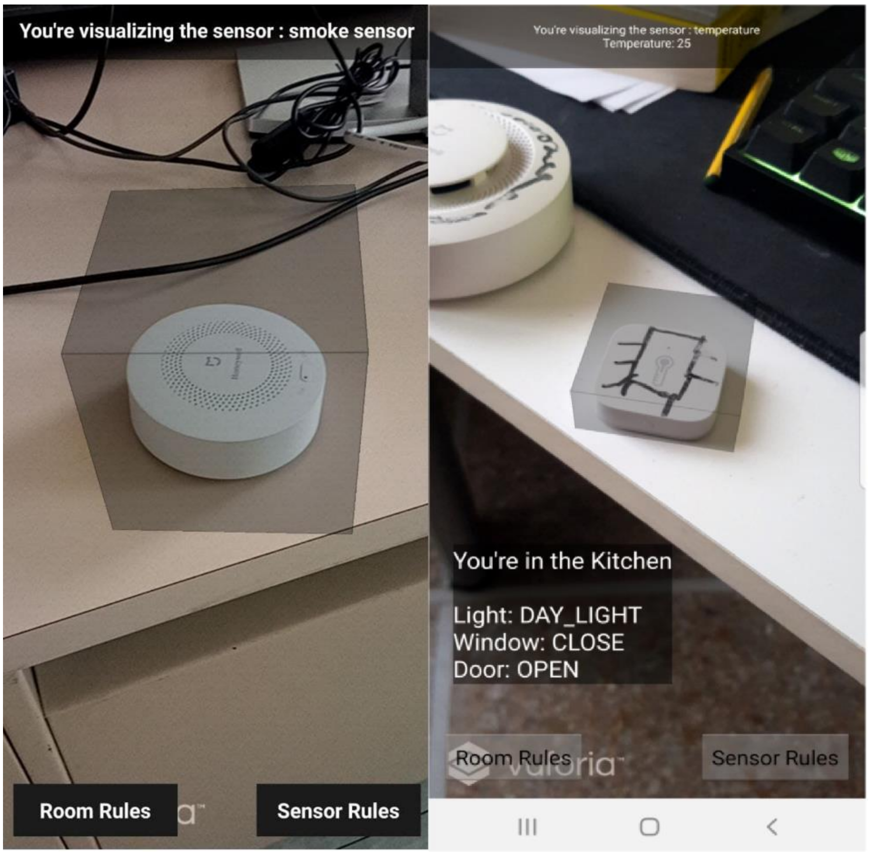

Figure 1: On the left, an object (Honeywell sensor) is recognized using Vuforia. The AR framework can place a 3d visualization over the image, in this example, a cube. On the right, the recognition of a temperature sensor (with a symbol drawn on it) cause the loading of related information.

[2] https://developer.vuforia.com/
[3] https://library.vuforia.com/features/objects/object-reco.html

3. Machine learning-based solution

Object detection is a computer vision technique, which has undergone significant improvement over the last years with the development of Convolutional Neural Networks (CNN). Object detectors rely on an architecture based on these networks to estimate the position and class of the various objects present in an image. We aimed to allow for detection using standard mobile devices (e.g. smartphones), which, as such, may have hardware constraints (e.g. in terms of battery consumption, computing power, etc.). However, state-of-the-art detectors are becoming more expensive due to the increased computational cost to execute the larger models [11]. For this reason, we tested EfficientDet, a family of object detector models able to achieve high accuracy and efficiency even when deployed on limited hardware. After tests on a Xiaomi Redmi 9 (a mid-range device released in 2020), we chose to carry out the training using version 2 of the model, because it offered a good balance between training time, accuracy, and detection speed. We constructed a prototype dataset of some common IoT devices and house objects, by downloading annotated pictures from OpenImages[4], and taking photos and annotating them using LabelImg[5] for objects not present in the aforementioned source. The downloaded training images were limited to 500 per object, while the training photos shooted were less than 150 per object. The training lasts around 5 hours on a pc with an Nvidia RTX graphic card, with a resulting average precision (COCO version) of 0.65 (between 0.79 and 0.88 on the objects with captured photos, lower on the OpenImages pictures).

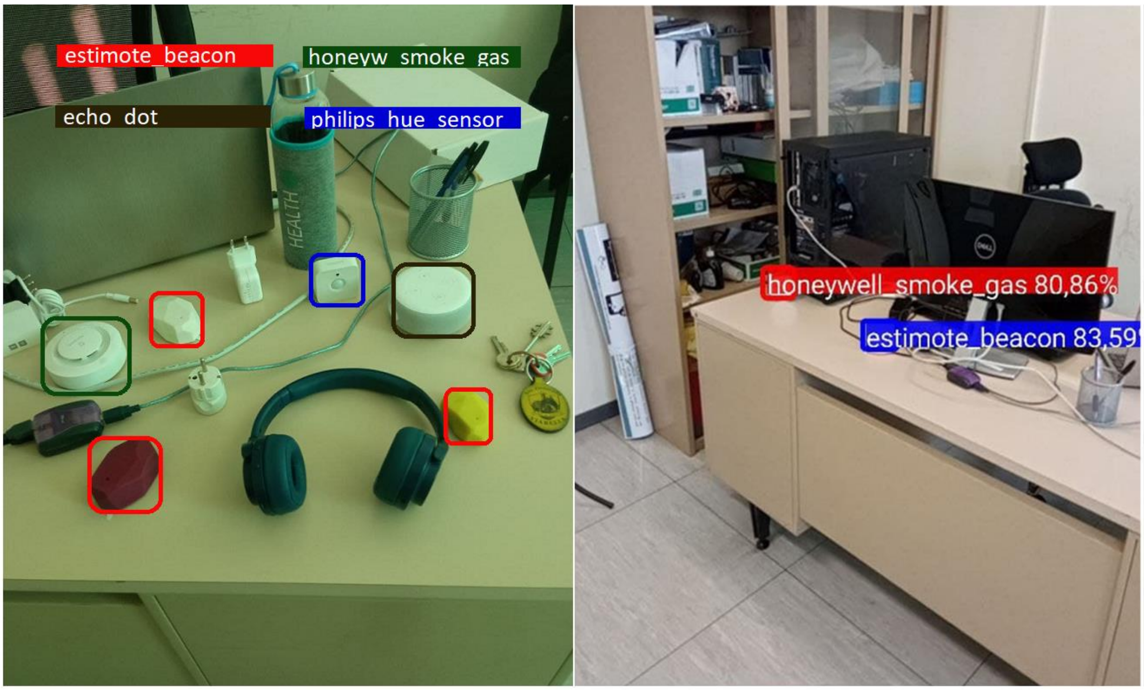

Figure 2: Objects detected using the neural network approach: on the left side, a crowded scene with many objects (labels edited for readability); on the right side, a detection from a longer distance.

4. Comparison between the two methods

From using the two methods for recognising objects, it emerges that each one has its peculiarities. The Vuforia based object tracking (see Figure 1) allows for a stable and accurate recognition, but it needs to get near to the object (around 50 centimetres or less for a small object like the Honeywell sensor shown in Figure 1) and focus the camera on it for some moments. However, to identify objects

[4] https://storage.googleapis.com/openimages/web/index.html
[5] https://github.com/tzutalin/labelImg

which have few distinctive features, there may be the need to draw some unique symbols or texture (markers) on them, which is not ideal because the recognition cannot be applied to other objects of the same type that do not present the same marker. However, if we have to consider a few objects and they have a distinctive shape, using the object scanner may be the preferred solution since the labelling and configuration phases do not require a lot of images, time and resources to train the recognition engine. Furthermore, AR frameworks provide techniques to integrate a 3D visualisation in real space, such as the capturing of the device position and orientation in space, the environment coordinates, and the collecting of feature points in the environment.

Machine learning-based detection (see Figure 2) has a fast recognition time (even if a small delay exists between framing and detection, the interaction is overall fluid), can identify multiple objects on the screen, and the classification quality is acceptable compared to the limited number of images used for the training. With the tested model and dataset, small objects like the Honeywell smoke sensor are detected up to around 150 centimetres, while for larger objects, this distance can be greater. This approach allows the abstraction from a specific model/device, allowing for reasoning based on classes of devices (i.e., a model can be trained with images of all versions of a device, which may have different sizes and shapes, but can still be labelled with the same class). In some cases, this may also be a disadvantage because, for example, all lamps will be recognised by the detector, regardless of whether they have smart capabilities. However, this problem can be easily solved using a "post-filter" on the detected object, removing the ones we know not to have such capabilities, based on the user location. This approach also has some drawbacks. The model sometimes stumbles upon false positives, but this problem can be limited by increasing the confidence threshold and eventually training a disambiguation class with objects similar to the ones that cause misclassification [10]. Also, even if large datasets of annotated images exist, to detect objects that are not present, there is the need to collect and annotate them. Finally, there is the need to fine-tune the model parameters to obtain a satisfactory performance/accuracy balance for the specific use case and avoid overfitting.

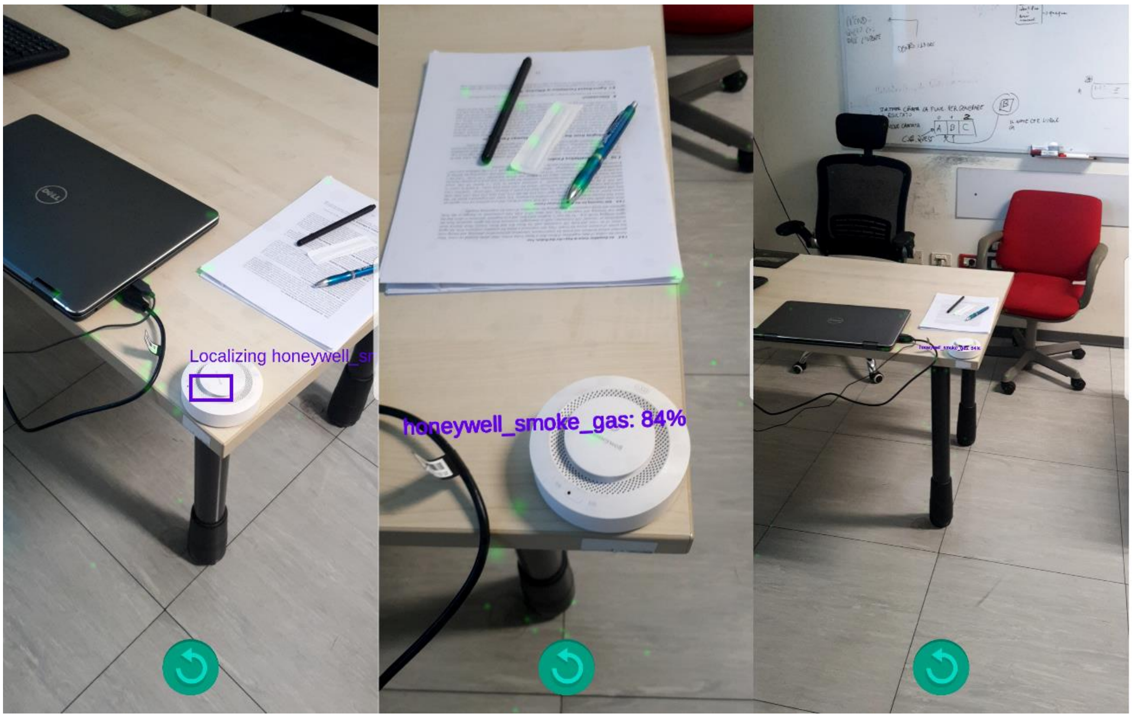

Figure 3: The machine learning-based detection (left side) is used as a reference to place the anchor, which will then be linked to that location in the environment (centre and right side).

From what was observed testing the two methods, a solution that uses the neural network object detection together with the capabilities of the AR seems promising (see Figure 3). From the user perspective, a tailoring environment that allows for immediate detection of the objects of interest (i.e., without the need to go near the object and start the detection) enables a more opportunistic exploration of the possibilities of the technology. Also, the possibility of framing more objects in the same view can be useful in the rule creation phase because it facilitates multiple triggers/actions composition by immediately selecting several recognised objects. Another possibility is using physical objects as proxies to render external data sources or concepts not directly generated by such specific objects. For example, framing a bed can enable the definition of a rule involving the sensor (undetectable by the camera) able to provide information on sleep time and quality.

5. Smart Home Scenarios

We envision three possible use cases for a smart home MAR application that uses object detection:

- Explorative rule visualisation and creation: selecting the items to be used in a rule from a list of possibilities is time-consuming [5] and can be a source of errors [4], especially when there are many options organised in a complex hierarchical structure. An AR-based approach, such as framing devices and selecting them using a tap, or drawing a connection between them, can be more immediate, intuitive, and less prone to errors to select them and to visualise their current state. An overlay can provide a summary of the rules involving an environment or a device, making it immediately visible if and when some objects are used in many rules (which may lead to conflictual or duplicate behaviours), or not at all. Connections between smart objects (e.g., rules involving multiple objects) can be visualised using various types of graphical representations.

- Augmented reality simulator: an AR-based simulator could be used to test from inside the environment the events and conditions that lead to the activation of a rule. For example, one or more rules could be selected from the rule repository and mapped to the environment to visualise if they would activate under the actual situation. Users should have the possibility to edit the current environment values, by acting on the augmented representation of the devices, and the simulator could also render the effects of the rule activation or trigger their activation. This "augmented debugging" seems particularly helpful to cope with the most nuanced aspects of personalisation (e.g., the difference between events and conditions, which may be rendered using different visualisations).

- Recommendations about possible automations: users expect that augmented objects can provide context-aware recommendations [6]. Suggestions could be obtained from the current state of the environment and/or from the analysis of the past data, used with the object that the user is currently framing (or the last-n framed ones), and be rendered as wirings between the involved objects.

6. Conclusions

We examined two possible approaches for recognising objects in an AR application for smart homes. Each one has its advantages and disadvantages, and the choice of one over the other may be case-specific. Nevertheless, the idea of using AR to customise a smart home environment is interesting and seems to open many valuable possibilities. As future work, we will design and implement some of the mentioned possibilities and evaluate them with user tests to gain insight into their usability and usefulness. Also, this type of solution can be compared with non-AR tailoring environments to understand the specificity of this approach better.

7. Acknowledgements

This work has been supported by the PRIN 2017 “EMPATHY: Empowering People in Dealing with Internet of Things Ecosystems”, www.empathy-project.eu/.

8. References

[1] Ariano, Raffaele, Marco Manca, Fabio Paternò, and Carmen Santoro, Smartphone-based Augmented Reality for End-User Creation of Home Automations, Behaviour & Information Technology, 2021, Taylor and Francis, London.
[2] Chatzopoulos, Dimitris, Carlos Bermejo, Zhanpeng Huang, and Pan Hui, Mobile augmented reality survey: From where we are to where we go, Ieee Access, vol. 5 (2017), pp. 6917-6950. doi: 10.1109/ACCESS.2017.2698164.
[3] Eloy, Sara, Luís Dias, Lázaro Ourique, and Miguel Sales Dias, Home Mobility Hazards Detected via Object Recognition in Augmented Reality, in: Proceedings of the eCAADe SIGraDi 2019 Conference, Porto, Portugal, volume 2, ISBN 978-94-91207-18-1.
[4] Gallo, Simone, Marco Manca, Andrea Mattioli, Fabio Paternò, Carmen Santoro, Comparative Analysis of Composition Paradigms for Personalization Rules in IoT Settings, in: Proceedings of the 8th. International Symposium on End User Development, IS-EUD 2021, Lecture Notes in Computer Science, vol 12724, Springer, Cham, 2021. doi: 10.1007/978-3-030-79840-6_4.
[5] Ghiani, Giuseppe, Marco Manca, Fabio Paternò, and Carmen Santoro, Personalisation of context-dependent applications through trigger-action rules, ACM Transactions on Computer-Human Interaction, TOCHI, 24, no. 2 (2017), 1-33. doi: 10.1145/3057861.
[6] Knierim, Pascal, Paweł W. Woźniak, Yomna Abdelrahman, and Albrecht Schmidt, Exploring the potential of augmented reality in domestic environments, in: Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '19, Association for Computing Machinery, New York, NY, USA, Article 31, 1–12, 2019. doi: 10.1145/3338286.3340142.
[7] Pawade, Dipti, Avani Sakhapara, Maheshwar Mundhe, Aniruddha Kamath, and Devansh Dave, Augmented Reality Based Campus Guide Application Using Feature Points Object Detection, International Journal of Information Technology and Computer Science, Mecs Publishing, Hong Kong, 10, no. 5 (2018), 76-85. doi:10.5815/IJITCS.2018.05.08.
[8] Platte, Benny, Anett Platte, Christian Roschke, Rico Thomanek, Thony Rolletschke, Frank Zimmer, and Marc Ritter, Immersive Language Exploration with Object Recognition and Augmented Reality, in: Proceedings of the 12th Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, pp. 356-362, 2020.
[9] Seiger, Ronny, Romina Kühn, Mandy Korzetz, and Uwe Aßmann, HoloFlows: modelling of processes for the Internet of Things in mixed reality, Software and Systems Modeling, 20 (2021), 1465–1489. doi: 10.1007/s10270-020-00859-6.
[10] Svensson, Jan and Jonatan Atles, Object Detection in Augmented Reality, Master Thesis, Lund University, Lund, Sweden, 2018.
[11] Tan, Mingxing, Ruoming Pang, and Quoc V. Le, Efficientdet: Scalable and efficient object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2020, pp. 10781-10790.