Design and Evaluation of an AR Voice-based Indoor UAV Assistant for Smart Home Scenarios

Robin Helmert
Bielefeld University
Bielefeld, Germany
robin.helmert@uni-bielefeld.de

Tobias Hardes
Paderborn University
Paderborn, Germany
tobias.hardes@upb.de

Enes Yigitbas
Paderborn University
Paderborn, Germany
enes.yigitbas@upb.de

ABSTRACT

As technology advances, Unmanned Aerial Vehicles (UAVs) have emerged as an innovative solution to a variety of problems in many fields. Automated control of UAVs is most common in large area operations, but they may also increase the versatility of smart home compositions by acting as a physical helper. For example, a voice-controlled UAV could act as an intelligent aerial assistant that can be seamlessly integrated into smart home systems. In this paper, we present a novel Augmented Reality (AR)-based UAV control that provides high-level control over a UAV by automating common UAV missions. In our work, we enable users to operate a small UAV hands-free using only a small set of voice commands. To help users identify the targets, and to understand the UAV's intentions, targets within the user's field of vision are highlighted in an AR interface. We evaluate our approach in a user study (n=26) regarding usability, physical and mental demand, as well as a focus on the users' preferences. Our study showed that the use of the proposed control was not only accepted, but some users stated that they would use such a system at home to help with some tasks at home.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**; **Usability testing**; **Sound-based input / output**.

KEYWORDS

UAV control, AR assistance, Usability Test

ACM Reference Format:
Robin Helmert, Tobias Hardes, and Enes Yigitbas. 2023. Design and Evaluation of an AR Voice-based Indoor UAV Assistant for Smart Home Scenarios. In *The 2023 ACM Symposium on Spatial User Interaction (SUI '23), October 13–15, 2023, Sydney, NSW, Australia.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3607822.3614522

1 INTRODUCTION

The idea of having intelligent assistants has long been a fascinating topic among people. For personal applications, these assistants can be used, for example, to control various Internet of Things (IoT) devices. In this manner, these devices can be programmed to work together and create a smart home [16]. There are many devices of this kind, such as switches, lamps, ovens, fridges, and smartphones. For instance, one can program switches or light bulbs to turn on the light when the user enters the room. A person is also able to program their oven with a smartphone and receive push notifications when it is preheated or when the meal is ready. These are just a few applications of what can be done.

The use of Conversational User Interfaces (CUIs) [35] or the Conversational Agents (CAs) [31] enables electronic agents to interact with an existing smart home remotely. Amazon Alexa, Apple Siri, and Google Assistant are prominent examples of such artificial intelligent agents. Typically, these interactions with smart homes are basic and do not automate workflows, but they do have an immediate purpose and often affect fixed devices like ovens, audio systems, or switches.

UAVs are one type of device that is rarely utilized in the context of IoT, smart homes, or CUI. Recent developments in the field of Human-Drone Interaction (HDI) have led to a wide range of use cases and application domains for UAV. A recent survey [17] shows that UAVs are being used in various domains, ranging from professional environments [24] to leisurely activities [7]. Here, UAVs can be operated either by an individual [30] or semi- or fully autonomous ([5], [22]), and can interact with users [8] or bystanders [36] alike.

The use and interaction of UAVs as personal aides in close proximity to humans remain understudied, warranting further investigation in scientific domains, such as *Help/Assistance* and *Companionship* as pointed out by the survey [17]. Historically, most UAVs have been relatively large and cumbersome, posing a danger to operators when they lose control. Currently, UAVs are available that are compact and stable enough to minimize the risk to users, which, in turn, reduces the required control distance.

In most professional use cases, UAVs are either controlled by a base station or operate autonomously while maintaining a distance from humans. Therefore, direct human interactions and influences are generally minimal. In contrast, for personal use, most UAVs are manually operated. Mastering standard UAV controls for most tasks require training, a general affinity for technology, and fine motor skills. However, this is less relevant in the described home IoT scenario, where the focus is on automating or simplifying workflows.

Automated UAVs could serve as small helpers to monitor and physically interact with the environment, such as delivering small items. They could exceed typical IoT capabilities, as IoT commands typically don't require real-life physical interaction. Although it's possible to lock, open, or close a door, these actions can be simplified to on/off commands for the door's actuators. The ability to command a physical moving body, such as a robot or a UAV,

SUI'23, October 13–15, 2023, Sydney, AU

ACM ISBN 979-8-4007-0281-5/23/10…$15.00
https://doi.org/10.1145/3607822.3614522

would offer more diversity. Similar to the robot Amazon is currently developing[1], it could move around the house, record, and stream using its cameras, and serve as a motion sensor to enhance security at night.

Detecting and transporting small objects could be the most advantageous feature. Avoiding human errors would require a UAV that not only ensures the safety of users in case of failures but also relies on highly automated control. It is worth noting that more automation implies less control and more unknown internal processes. Moreover, if people do not comprehend what is happening, they feel insecure. This could also increase the UAV's fear due to fast-moving parts and uncertain movements.

One way to alleviate these fears is by providing safety measures, such as using drone cages or safety glasses. It is also important to consider the controls themselves. The controls should be intuitive and straightforward. UAVs and their control systems are widely discussed in current technological developments. For instance, Konstantoudakis et al. [25] conducted a study comparing different gesture controls. Erat et al. [12] developed an AR enhancement to control UAVs and explored hidden areas, while Huang et al. [20] propose a voice-based UAV control utilizing a full language model.

In this paper, we focus on the research of voice-controlled UAVs, which could act as a personal helper in smart homes in the future. We propose a prototype UAV control that focuses on semi-automatic target calculation, utilizing an AR head-mounted display for visual aids and providing an easy selection of flight destinations through voice commands. Since it does not require any prior UAV knowledge or fine motor skills, it makes UAV control more accessible and applicable in smart homes. The paper also addresses the issue of general fear of UAVs, particularly at close range, for this use case. The proposed control's ease of use should help establish the foundation for integrating UAV into smart homes as personal helpers.

We evaluate the efficacy of the proposed UAV control based on how well lay users can understand basic target detection and position calculation, and how likely they are to use a UAV assistant at home if one were available. To be more specific, our study aims to ascertain if lay users can use a UAV with the provided features in a satisfactory manner in an IoT environment. Various metrics like usability, physical or mental demand, and user preference are utilized to evaluate the usability of the proposed UAV control.

2 RELATED WORK

As the popularity of IoT and smart homes increases, potential technologies that can be integrated are being researched. The following sections discuss the literature on smart home applications, CAs, and UAV interactions, in that order.

2.1 Augmented Reality

Augmented reality is a technology that enhances reality with optical effects for various purposes. Various AR technologies exist, employing diverse techniques and immersion levels. The technology used in our work is the Microsoft HoloLens 2, a head-mounted display with see-through glasses. These glasses can project stereoscopic images to create three-dimensional objects in the real world. The HoloLens 2 can track its position and orientation, making it possible to create anchored holograms in the real world. The HoloLens 2, with its form of augmented reality, is applied in various fields, such as electrical engineering [43], and can also be utilized for computer vision [44]. In cases such as surgeries [39], the possibility to place and interact with stable holograms while keeping the hands sterile enhances the information flow significantly.

The HoloLens 2 is not yet widely adopted for personal use due to its relatively high cost. Nevertheless, technological advancements may make it more affordable, thus opening up possibilities for meaningful use cases such as smart home integration.

[1]https://www.amazon.com/Introducing-Amazon-Astro/dp/B078NSDFSB

2.2 Smart Home Applications

Recently, there has been a growing interest in smart home applications, as demonstrated by the increasing number of research and survey papers published, such as those by Stojkoska et al. [42], Gunge et al. [15], and Marikyan et al. [33].

De Silva et al. [10] also conducted a review of state-of-the-art smart home technologies, with a particular emphasis on audio and vision-based techniques. For instance, one of the papers reviewed in their study used computer vision to recognize human actions such as falling, walking, or standing [9]. When used in combination with audio-based systems, it is possible to recognize other activities such as coughing and closing a door. Both systems operate as sensors that identify specific situations, allowing them to respond to these situations. They are not intended for active control.

Overall, there has been very limited research regarding the integration of UAVs into a smart home environment. For instance Xia et al. [47] furnish a brief demonstration of UAVs in smart homes. Hence, this paper aims to establish the fundamental principles for the integration of UAVs into smart homes by providing a suitable drone command strategy. As a general trend, smart home environments frequently support the use of CAs to interact with the users and integrated devices.

2.3 Conversational Agent

According to a literature review by Mariani et al. [32], the use of conversational agents in smart homes has substantially increased.

Diederich et al. [11] conducted an extensive review that analyzed different interactions with conversational agents, including enhancing user request efficiency and exploring the role of conversational agents as car assistants [13, 28].

Moreover, the review placed focus on studies that examined automation in organizational processes, including customer service [1] and sales [45]. While all the mentioned conversational agent applications aim to automate or speed up processes, they can only provide digital actions. In contrast, we provide the groundwork for a mobile UAV conversational agent that can fly around and, in the future, interact with the environment.

Sciuto et al. [41] conducted another study on CAs for personal benefits. The authors studied how Amazon's conversational agent, Alexa, is implemented and utilized within households. They discovered that the most typical commands for Alexa are basic, such as 'Alexa, what time is it?', 'Alexa, what's the weather like?', 'Alexa, tell me a joke', or 'Alexa, turn on my light.' These interactions ease the user's workload but are restricted to the devices integrated into the smart home system. For example, Alexa is unable to control a lightbulb or switch that is not linked to the smart home network. UAV helpers can extend applications by interacting with non-connected items, despite their small size. Future versions could turn on lights by physically pressing a button, without requiring a connection to the bulb or switch. Only the location is important, and it can be effortlessly detected using the presented setup. However, operating and commanding a UAV is not always an easy task to learn or do.

2.4 UAV Interaction

In general, there are various interventions involving UAVs with advantages and disadvantages for the user or the UAV's versatility. Herdel et al. [17] provide a recent survey that reviews and classifies UAV research and UAV application domains.

Recently, research has focused on tangible human-UAV interactions. In their study, Huang et al. [20] suggested utilizing voice-controlled commands for navigating a UAV within a grid-based environment measuring 4x4x2. It employs a language solver to understand more complex sentences instead of basic commands. To interpret complex sentences rather than mere instructions, a natural language processing (NLP) algorithm is employed. Thus, users can control the UAV with their own vernacular without the need for memorizing specific instructions. Nevertheless, the UAV's capabilities are limited despite these advancements in control technology. The UAV can only operate in a confined space that includes thirty-two possible positions defined as 4x4x2. This approach is not suitable for a household where objects and targets can vary and are not consistently uniform. Unlike a grid-based system, the UAV control presented here operates independently. The UAV can reach any detectable object within its range.

In their research, Erat et al. [12] use Microsoft's HoloLens to maneuver a UAV while Line Of Sight (LOS) is obstructed by obstacles, such as walls. They study how effectively users can manipulate a UAV even when the line of sight is obstructed. To do so, they compare three distinct control systems. In the initial control strategy, a joypad is employed to control the steering of the UAV, which itself is not visible. Instead, a live feed from the UAV and an up-to-date 3D visualization of the UAV's position within the room, along with its interiors, is displayed on the monitor screen in real-time. The second control named *pick and place* uses gestures to move a virtual representation of a UAV. The real UAV follows the movement afterward while avoiding obstacles. To enhance the control the virtual walls of the obstructed room are displayed in AR on a HoloLens. The third control is similar to the second, but the UAV's movement uses gaze as direction determination.

Another AR-based approach was done by Konstantoudakis et al.[25]. They present a UAV tracking system that combines visual detection with cumulative Inertia Measurement Unit (IMU) data to estimate the UAV's position. They have implemented two gesture-based UAV controls. For both, they used a HoloLens 2 headset for AR-enhanced feedback and gesture-tracking purposes. The first control is palm-based where a UAV is controlled by adjusting one's palm orientation. For example, the UAV is commanded to fly straight ahead by tilting the hand forward and back by tilting it backward. The second control is similar to gesture-based but does not consider the palm orientation and movement but rather the finger movement.

Although Erat et al. [12] and Konstantoudakis et al.[25] enhance their UAV interactions with augmented reality, they stay with a manual UAV control, resulting in an immediate low-level UAV response for each user action. In contrast, UAVs in our work are controlled on a high level, meaning it is sufficient to select a target while other tasks like path calculation are done automatically.

Peining et al. [37] present a more exotic approach to interact with UAVs using a hands-free approach that is based on Brain-Computer Interface (BCI) data. They compare an Emotiv Insight and a Muse 2014 EEG headset for their accuracy and usability for controlling a small UAV based on brain activity. Besides the LOS, no additional enhancements are given. Since the controls with such EEG headsets accept only concentration and relaxation as inputs, they are limited to one-dimensional actions, and therefore only a very rudimentary UAV control can be created.

There are also other UAV controls, which are not intended for smart home applications directly, but for smart cities, smart agriculture, and others. They are created for the same purpose, simplifying and automating workflows. For example, UAVs are used in agriculture to measure soil pollution [21], monitor vineyards [40], or detect stinted growths [19]. Some UAVs also utilize deep learning on images to detect objects [6], avoid obstacles in indoor races [23], or early detect sinkholes using thermal cameras [29]. However, most of them use image processing to either find objects in the images or avoid obstacles. None of them deal with UAV controls that actually use the resulting information to define targets the UAV should fly to. Furthermore, there is research where UAVs can automatically detect targets even with spatial mapping. Bergé et al. [3] present a point cloud-based spatial room scanning method for UAVs to detect possible targets and Boudjit et al. [4] researched target detection with QR-codes. However, none of them researched how those target detection results can be utilized to develop an actual UAV control that is intuitive and may be usable in a smart home like the prototype presented in this work.

3 CONCEPT AND IMPLEMENTATION

In this section, we first provide an overview of the solution, describing the general workflow. Subsequently, we delve into a more detailed explanation of the implementation and the workings of the UAV control.

3.1 Solution Overview

The overall aim of the prototype designed here is to simplify the control of a UAV such that it can be used by users not used to control UAVs. Therefore, the control itself is reduced to voice commands to instruct the UAV to valid targets. Valid targets are hereby calculated and highlighted with the aid of Microsoft's HoloLens 2. The highlighting of targets is expected to increase transparency to the user by helping to understand where and how the UAV will fly.

We have chosen a Crazyflie 2 [14] equipped with a lighthouse positioning deck as a UAV. The lighthouse deck is essential because it provides a stable coordinate system given being in the range

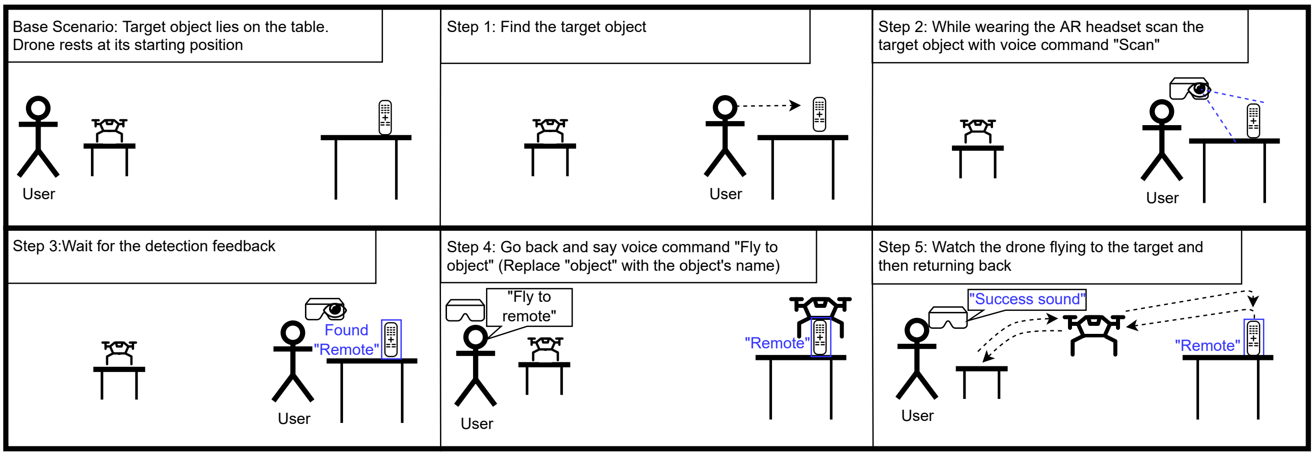

Figure 1: Example use case scenario for an intelligent aerial assistant. The manual scanning step may be removed in the future for even more automation.

of SteamVR Base Stations 2.0[2] and therefore can accept precise position commands.

The general workflow is as follows. The calculation of target coordinates is done by utilizing a HoloLens 2 with its spatial mapping functionality. It creates a wireframe environment model and thus a 3D model of the surroundings including a stationary coordinate system. With photos made by the HoloLens, the position where they were made, an object detection algorithm, and the environment model it is possible to track the virtual positions of actual objects. When the coordinate systems of the HoloLens and the UAV are synced, and the UAV is commanded to such a virtual object position, the UAV will automatically approach the real object. Users can do this by voice commanding the UAV to the object that initiated the position calculation. An illustration of the scenario is shown in Figure 1. Users can scan objects by watching them and invoking the scan command. Afterward, the field of view of the user is scanned for objects. When one is found it is highlighted in AR. This procedure allows as many objects to be highlighted and flown to as found in the scanned image as long as the object is visible to the cloud detection algorithm. Finally, the user can command the UAV to approach a target by saying fly to *object*, where *object* is replaced by the actual name of the target found.

3.2 Implementation

Beginning, with the engine, Unity[3] was chosen to run on the HoloLens 2 for the calculations, and Python ZeroMQ [18] was used on a notebook acting as a server for the communications between the UAV and the HoloLens 2. Therefore the setup consists of the HoloLens 2 (in the following called just HoloLens), a server running on a notebook or workstation, and the UAV as the main components. In the following the main workflow and implementation concepts are explained in more detail.

[2]https://www.vive.com/us/support/vive-pro/category_howto/about-the-base-stations.html (Accessed: 21.06.2023)
[3]Unity Repo: https://github.com/AnonymousGit2/UAV_Assistant_

3.2.1 Foundations. At the start, without needing any user input, the HoloLens makes an automatic spatial mapping of the room without any notations of where which object. As a result, a 3D model of the environment is made which then can be used as the foundation of the subsequent interactions and calculations. To make an object known to HoloLens, it has to be scanned with the built-in camera. In the following, a more detailed description of the workflow for obtaining the 3D coordinates of the target object from a single HoloLens photo is explained. For better comprehension, the different steps explained in this chapter are also shown in Figure 2.

3.2.2 Scan Workflow. The *Scan Workflow* starts with the *Scan (1)* command, which forces the HoloLens to take a photo (2). When the HoloLens takes any photo, a *PhotoCapture (3)* object is created, containing, amongst others, the image *Texture (4)* and a *CameraMatrix (5)*. The *CameraMatrix (5)* contains the camera's position and projection matrix; both are important for further calculations. The image *Texture (4)* is sent to a cloud vision service (6) to detect the objects in the photo. For our prototype, we have chosen Google's cloud vision service but in general any other with object detection would also work. After the image was sent, the service returns the names of the objects found and their bounding boxes (7). The bounding boxes are returned in the form of four coordinates describing the corners of the rectangle. The coordinates given are always relative percentages with their origin in the top left corner. In parallel to the action above, the Unity application calculates a *BlendedImage (8)*. A *BlendedImage (8)* is a spatially static picture hologram that, when viewed from the position it was taken, is indistinguishable from the actual environment behind it and therefore blends with its surroundings, and therefore, the *real object* is directly behind the object in the *BlendedImage (8)*. Such a *BlendedImage (8)* is calculated using the camera position and the projection matrix, which were both saved when the photo was captured.

Based on the position where the photo was taken and the spatial mapping of the real world it can be used to calculate the positions of real objects as illustrated in Figure 4.

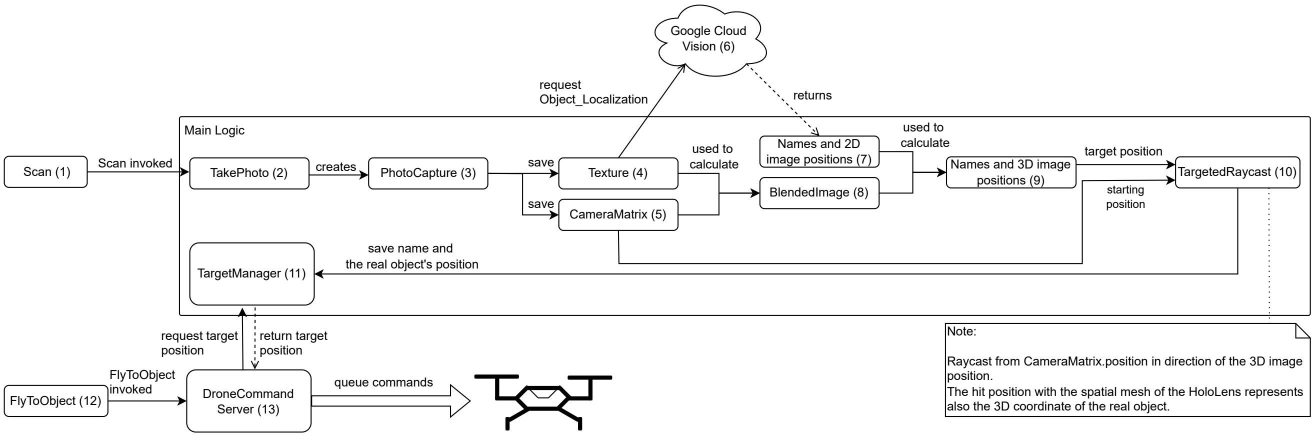

Figure 2: Workflow from giving speech inputs to the UAV approaching the chosen destination.

The mapping shows that the extended line from the position the photo was taken through the blended image would hit the real object at exactly the same position and with the spatial mapping provided by the HoloLens exactly this is possible. Here raycast can be cast from the position the photo was taken in direction of the object's position in the blended image to hit the spatial model at the position of the real object. This position can then be used to augment the real object or to directly command the UAV to fly to that position.

What remains is to calculate the direction of the raycast used in 3D space. Since raycasts need 3D coordinates, but pictures are only in a 2D space, the coordinates of the found objects in the image need to be transformed first (9).

A schematic transformation of the relative 2D image coordinates into 3D Unity coordinates can be seen in Figure 3.

Here the blue Arrow is calculated as follows:

(1) The values *image width*, *image height*, and the *image position* -in the following called v_{origin}- as well as the *image orientation* are given in Unity.
(2) Now start with the *image position* vector v_{origin}.
(3) Calculate the normalized vector direction along the x-axis of the picture in 3D space v_{xnorm} utilizing the *image position* and the *image orientation*.
(4) Calculate the needed length by multiplying the x-value of the *target position* with the *image width* w_i. Together with the normalized direction v_{xnorm}, it results in the 3D vector $v_x = v_{xnorm} \cdot w_i \cdot 0.75$.
(5) Calculate the normalized vector direction along the y-axis of the picture in 3D space v_{ynorm} utilizing the *image position* and the *image orientation*.
(6) Calculate the needed length by multiplying the y-value of the *target position* with the *image height* h_i. Together with the normalized direction v_{ynorm}, it results in the 3D vector $v_y = v_{ynorm} \cdot h_i \cdot 0.25$.
(7) By adding the three vectors the final vector $v_{pos} = v_{origin} + v_x + v_y$ is derived. It is only correct for that image, and only as long it is not moved or rotated after the calculation. For each additional image or position, the procedure has to be repeated.

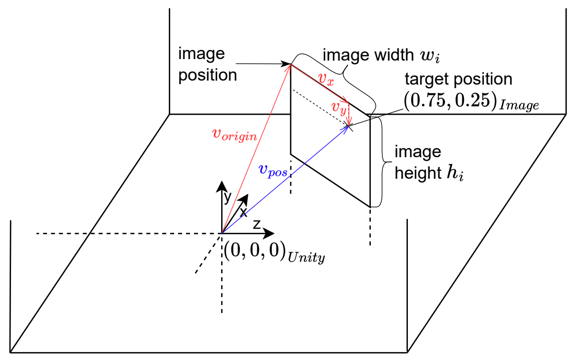

Figure 3: Illustration on how 3D coordinates can be extracted from 2D image coordinates in Unity

During the application execution, the *BlendedImage* is made invisible since it is only essential for the calculations and not for the users to see.

Since the resulting coordinate can be used to instantiate the raycast (10) and thus to get the position of the actual object it is now possible to calculate and save the actual positions of all photographed objects which the cloud algorithm can recognize to the *TargetManager (11)*. The *TargetManager (11)* acts as a Database where the positions of all found objects are stored and updated. By converting that position from the coordinate system of the HoloLens to the one of the UAV it is possible to calculate positions the UAV can understand and therefore approach. It then remains to command the UAV to that position.

3.2.3 Commanding the UAV. Afterward, the UAV can be instructed with the *FlyToObject (12)* command where the phrase *Object* can be replaced by any of the found objects. The *DroneCommandServer (13)* will then request the current position of the defined object from the *TargetManager (11)*, calculate a primitive path to that position,

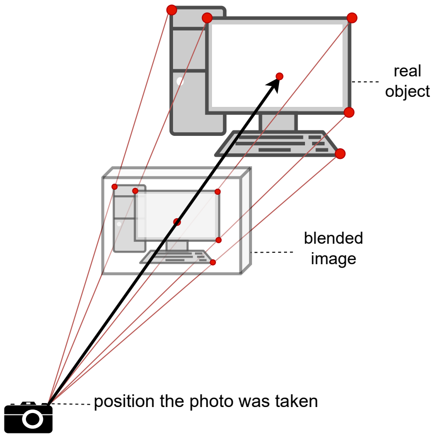

Figure 4: Workflow of generating the final target positions from the 3D image coordinate calculated before

and finally queue the resulting positions commands to the UAV. In our current version, obstacles are not avoided, as the UAV does not have the measurements to do so. However, obstacle avoidance for UAVs is already implemented by other researchers like Woods et al. [46] or McGuire et al. [34].

In the following, the field usage of the described calculation is illustrated. To fly the UAV to a target position, users have to walk to the target, look at it and say *scan*. The HoloLens automatically scans the object and then augments the users' vision with an object marker at the position of the real object. Figure 5 shows the resulting marker from the scan operation.

Next to the marker, a tooltip showing the object's name is also presented to the users. In the case of the figure, it reads "Command to fly here: 'Fly to Computerkeyboard'" as the computer keyboard was the name of the object returned by the cloud vision algorithm. With the tooltip, users can directly see the command that lets the UAV fly to that object. To finish the task, it remains to command the UAV with the voice command *Fly to Object*, where the actual object's name replaces the object phrase. In the case of Figure 5, this command is *Fly to Computerkeyboard*. Then the flight path is calculated, sent to the server, and ultimately to the UAV. The UAV then flies along the calculated path and returns to the starting position afterward.

4 EVALUATION

We conducted a user study to evaluate the proposed UAV control system.

4.1 Hypothesis

This study's objective is to evaluate whether everyday users can adequately operate a UAV in an IoT environment, along with their readiness to embrace and utilize it. In order to accomplish this, we investigate the following hypotheses.

Our first hypothesis states that the UAV's control system is highly usable and automated, leading to a high System Usability Scale (SUS) score.

Our second hypothesis examines whether the workload involved in operating the UAV is low, indicated by the combined NASA Task Load Index (TLX) score, along with individual scores related to physical and mental demands.

Our third hypothesis predicts that the UAV's small size will alleviate any concerns or fears among users, encouraging them to use it in real-life situations.

4.2 Procedure

We began by welcoming the participants and providing them with an overview of the topic, as well as explaining safety measures and data privacy. This introductory phase lasted about 15 minutes. Following that, we provided the participants with roughly 10 minutes to answer some basic questions about themselves. Next, they were given a brief amount of time to prepare by solving the tutorial we provided. The tutorial was focused on the use of augmented reality (AR) to provide an explanation of how things would work. Following the tutorial, the participants were required to complete the main task.

The primary objective was to fly the UAV to a designated target and then return to the starting position. Users had to identify the targeted object and learn the necessary procedures to pilot the UAV to the objective. The tasks were not specific for the target object and could have been replaced by any other object.

However, since the proposed voice control relies on scanning and detecting an object, it is crucial that the detection algorithm can recognize that object. Changing the objective introduces the problem of various tracking accuracies of different objects. While the cloud algorithm can detect many objects, some are difficult to determine based on their general appearance alone.

The target object remained the same during the experiments because the ease of detecting it directly impacts its usability and mental demand, which rely on the cloud vision service rather than the control strategy. This ensured that all users had the same conditions. It was assumed that even if the target differed among participants — whether they had to guide the UAV towards a pen or remote control — their usability and workload would remain unchanged. (System Usability Scale) (SUS) and TLX questionnaires were distributed to participants upon task completion.

4.3 Participants

Our study involved 26 participants, ranging in age from 22 to 32 years, with 6 women and 20 men included. Of these participants, 23 were pursuing bachelor's or master's degrees in computer science, while the remaining three were studying media studies, taxation, and philosophy. Given the predominance of computer science students, the participants generally displayed a notable inclination towards technical matters. Specifically, 18 participants rated their technical affinity as *high*, four as *very high*, and four as *average*, with no one indicating low or very low affinity.

When we asked people about their experience with augmented reality (AR), we found that most of them were familiar with using AR on smartphones, while their experience with AR on Head Mounted Displays (HMDs) was relatively limited. Taking into account the proficiency of the participants with smartphone-based AR, 12 rated themselves as *high or better*, five as average, and nine

Figure 5: Illustration of the highlighting of targets done in augmented reality. After using the command *Fly to Computerkeyboard*, the UAV approaches the target according to the calculated path.

as *low or less*. Regarding HMD-based AR, 6 participants rated themselves *high or better*, six as average, and 14 *low or less*.

Similarly, participants were asked to state their experience in controlling unmanned aerial vehicles (UAVs). Only two participants rated their UAV control experience as *high or better*, while two rated it as *average*. The remaining 22 participants reported *low or lesser*, indicating limited familiarity with controlling UAVs.

The last demographic question asked to the participants was if they fear UAVs at close distances. The question had four possible answers, *not at all*, *depends*, *kind of*, and *yes*. If *depends* was selected, then to get more insights, it was also asked on what it depends. Of the 26 participants, four were *kind of afraid* of UAVs, and eight selected *depends*. Most of the reasons for the *depends* answer were about the size of the UAV and whether a protective eyeglass is worn.

So the people stated felt safe if the UAV was small and safety glasses were available. Even though the experiments were designed utilizing a small UAV compared to other state-of-the-art ones, and the participants had to wear safety glasses, it was expected that they felt safe enough not to influence the outcome negatively.

4.4 Results

After the study was conducted, we evaluated the results based on the given SUS- and TLX- values that are presented in the following.

A typical SUS questionnaire consists of ten statements on usability. Users have the choice to mark one number ranging from 0, for strongly disagree to 5, for strongly agree. After collecting the data, the score was scaled and averaged over all participants. We calculated the mean SUS score of the UAV control with a value of 85.9. Bangor et al.[2] created an adjective rating to interpret these values, and according to that rating, this score is excellent and therefore the second-best option.

The other measure we investigated was the NASA TLX. It resembles the cognitive and physical demands of the participants during the conducted study. Like the SUS score, the values of the TLX are also classified as a Likert scale. Unlike the SUS score, lower TLX values indicate a better result. Higher values show therefore higher physical and mental workload. Users could here choose an answer between 0 and 20 stating the amount of workload they had. After scaling, the highest score possible is 100 and the lowest score is 0. For our prototype, the average TLX value lies at 20.83. According to the scale discussed by Prabaswari et al. [38], it resembles a medium general workload.

To gain further insight, the first two questions of the TLX questionnaire are evaluated separately with the objective of directly analyzing the perceived physical and mental demands. The physical demand score had an average of 17.12 and the mental demand an average of 31.06. Therefore, the physical demand is rated *medium*, and the mental demand was rated *high*. We suspect that the latter is caused by the short distances to the UAVs, but further research is needed to understand exactly how this affects the mental demands of the users. The lower physical load can likely be explained by the limited need for physical movements, whereas the heightened cognitive load may be due to the novice experience and proximate interactions with UAVs. However, further empirical research is needed to definitively assess the reason for the high mental demand.

4.5 Qualitative Results

In evaluating the results, it is imperative to consider not only the quantifiable metrics of usability and workload but also the subjective responses and reflections from the participants. Consequently, to capture these nuances, two supplementary questionnaires were administered post-study.

Feature Comparison. We used a first questionnaire to compare different features of the UAV control. It included the following questions (possible keyword answers are underlined):

- In general, would you prefer augmented reality enhanced vision during UAV flights, or would you rather see only the real UAV without any virtual enhancements? Maybe even other enhancements?
- Having a UAV helper at home, would you prefer to set its targets manually or have a set of possible targets acquired automatically?

The responses obtained from these inquiries indicated that out of 26 participants, 21 expressed a preference for vision enhanced by augmented reality as opposed to non-enhanced vision. Furthermore, in the scenario of having a UAV assistant at home, 14 of the 26 participants expressed favor for automatic target acquisition over manual methods.

Pro and Contra arguments. Subsequently, the participants were requested to enumerate arguments both for and against the subject under consideration. Similar comments were counted, and the most prominent points for each perspective are outlined in the following section. The numeric value enclosed in parentheses signifies the frequency with which each comment was expressed.

- Positive: Easy to use (13), Physical and mentally easy to use[5], Least amount of work as the UAV and the headset does everything (5), Intuitive (3)

- Negative: Flight path not clear (5), Disliked because voice commands are generally unreliable (4), Rely on good object tracking (2), UAV does not avoid obstacles (2)

The prototype was praised by the participants for its simplicity, intuition, and the little amount of work necessary to achieve the goal. However, it also had its downsides. We noticed that some participants were negatively biased regarding the concept of giving voice commands. They stated voice commands were in general unreliable. Others criticized the need for an Internet connection to the cloud vision service due to data privacy reasons. More prominent is, however, the argument given by several participants, that they did not know the flight path of the UAV. They wanted to know and configure it precisely and in general felt unsure about controlling the UAV, by not knowing the exact path it would fly.

Free Comments. Finally, the participants were asked to give free comments based on their experiences with the given UAV control. Some comments included expressed doubt in voice and object recognition. Others mention their favor for standard UAV controls since they are more fun to use. Some arguments against the control were caused by distrust of voice-based or cloud computing systems and their privacy policies since they do not explicitly know where their data is concretely sent to. These participants said they would use the voice control if it was completely offline and no private data could be leaked. Again, others simply stated that they liked controlling the UAV while experiencing AR and had a lot of fun doing so.

Finally, the users were asked two verbal questions. It was done verbally, so they would give an intuitive and direct answer without thinking about it.

First, participants were asked if they were afraid of the UAV during the study. Nearly all of them answered with a strict no and reasoned that the smallness of the UAV would not render a threat to them, which is an important result and the reason why we used a small UAV.

Second, participants were asked if the UAV could bring them objects hypothetically, would they use it to get such objects? Ten participants answered with a strict yes, and ten others stated that it would depend on the item's situation, size, value, distance, or the scenario in general. Only six gave no clear answer or answered with a no. Although the used UAV could not bring items, we added this question so that users could think of a use case where a smart home integrated UAV could be useful. Therefore, a UAV helper at home that can carry items could be advantageous in the future, as there exist users who would probably use such a helper.

UAV Performance. In addition to the evaluation given by the behavior and feedback of the participants, the performance of the UAV was also considered. However, due to the fact, that the workwise is relatively independent from the concrete hardware performance, this is only discussed shortly. While the UAV was approaching its targets the path it should fly was concretely calculated using position waypoints the UAV should fly along. During the flight the UAV was relatively stable in mid-air. Nevertheless, when slowing down, for example changing movement direction necessary when reaching a position waypoint, it jiggled around a bit. In general, when a target was found, the UAV managed to reach it every time within a two-centimeter accuracy, which good enough for the experiment but needs to be improved in the future.

4.6 Discussion

Our target was to answer the research question of whether lay users can use a UAV with the given features in a way that they would accept and use them in an IoT environment.

We started with the question of whether users would accept a UAV in close proximity. To answer this question, we requested them to state how scared they were of UAVs. We asked them twice, once before they did the experiment and interacted with the UAV, and once after the experiment was completed. The majority stated that their fear of UAVs depended on the size and whether protective eyeglasses were available or not. Since both reasons were covered with the small Crazyflie UAV and the protective eyewear, nearly none of the users stated they had any fear of the used UAV in the end. As a result, we deduct, that such a UAV would be accepted and used in an IoT environment when it is small enough and does not render a threat to the users. Since each participant had to wear protective eyeglasses, and the UAV was too small to cause injuries, these conditions were fulfilled during the experiments. However, it can also be solved by providing a UAV rotor cage as presented by Kornatowski et al. [27], [26]. Thus, the question of whether users would actually use a UAV in a smart home depends mainly on the concrete implementation of the features and their usefulness.

To measure the usability of the implementation features we used the System's Usability Scale and the NASA Task Load Index. Regarding the SUS score, and according to the adjective scale presented by Bangor et al. [2], the prototype was rated excellent.

Regarding The NASA TLX overall, the prototypes workload was medium. Comparing the mental and physical demands, the physical demand was lower than the mental one, which had been rated relatively high. The latter score might be biased by the fact that many participants worked the first time with a UAV and were afraid that they could break it. However, it has to be researched further with additional experiments to show whether this hypothesis is true or not. Regardless of whether this hypothesis is true or not, with a bit of training the mental demand might be lowered. Therefore, in relation to the research question, we conclude that the UAV and the given features can be used by lay users in a useful way.

The last question to be answered is whether the given features are useful enough to be used in an IoT environment. An example use-case in such an environment is, for example, carrying small items. The prototype itself did not realize a carry or bringing functions but rather the foundation to do so. Therefore, when referring to that use case in an IoT environment most participants stated they would make use of the system.

In summary, it can be said that the research question can be confirmed when considering the concrete prototype and the use case presented here. For additional use cases and features, the software, the UAV, and the interactions need to be extended and improved in some aspects.

4.6.1 Additional Remarks. In addition to the answers to the research question, other notable findings have been made. The findings made suggest a medium general workload and a higher mental demand. However, many participants stated that the system is easy to use requiring only a minimal amount of effort. This contrast shows a discrepancy between the results shown by the NASA-TLX values against the free-form answers given by the participants. This could be attributed to the relatively high technical affinity and background, such that the participants are used to working in AR or with UAVs. That might not really reduce the mental load but make it less stressful for them resulting in higher TLX-values while experiencing the process itself as easy to use. Another cause could be grounded in the presence of UAVs. Since participants stated that they are kind of afraid of UAVs under different conditions, they could continuously be aware of the UAV resulting in relatively high mental demand while the commanding process seemed to be easy and less demanding for them.

Although the study has provided valuable information and notable findings, it is important to acknowledge certain limitations that can impact the interpretation of the results.

5 LIMITATIONS

For the voice-based system, it was expected that some voice commands are not always recognized correctly. For example, people who have a heavy accent or are not used to the English language may be sometimes not recognized by voice detection algorithms. When that happened, they then would try to repeat the command but remain unsuccessful which increases frustration, the amount of errors done, and reduces self-confidence However, during the actual study, most commands got recognized, even if not, the participants were used to voice commands sometimes not recognized.

In contrast, errors in the automated target acquisition caused by scanning the environment were not understood so quickly. When invoking the scan command front camera of the HoloLens is used to take a picture. Therefore, head-shaking or other head movements can impact the detection negatively. If the detection algorithm gets a blurred image, it will be hard to detect the proper objects.

The most common error of image detection was that no object was found. Consequently, users got the feedback that no object was found and had to redo the scan procedure. During the study, no hints were given without an explicit question, and therefore such errors may have raised the measured mental demand.

During the study, 26 participants tested the UAV control. Most of them had a high technical affinity and had used AR devices before. For more representative results, the control needs to be tested more extensively with more participants and higher background diversity.

This work examines a generic control approach starting with a simple task. The setup chosen focused on performance and comparability, and therefore one task with only one and the same object was chosen for all participants. Whether multiple possible and changing targets affect usability and load was not considered for now and is needed to be researched in the future.

The control approach itself is also applicable to other scenarios, and the hardware is exchangeable. However, whether the same results can be concluded for tasks significantly different from those described in this work must be also researched further.

Additionally, although the UAV's positioning precision was satisfactory nearly all the time during the study, this aspect was not measured because the focus was on the control itself. Precision measures and related errors need to be elicited for further use cases. In addition to that, further tasks with more variety need to be tested for more meaningful results.

An important aspect to mention when working with UAVs is safety. During the study, the environment was fixed, users were not allowed to enter it while the UAV was flying, and they had to wear protective eyeglasses. This way it was guaranteed that the UAV could not render any harm to the users, even though it would be highly unlikely due to the small size of the UAV. Nevertheless, when considering real-life applications over multiple rooms with multiple users moving around safety needs to be guaranteed. One solution for this can be so-called UAV cages like those presented by Kornatowski et al. [27], [26], but which measures need to be additionally taken, needs to be further researched in the future.

6 SUMMARY AND FUTURE WORK

In this work, a voice-based UAV control that emphasizes automation was conceptualized, implemented, and evaluated for integration in a smart home environment. To command the UAV users have to invoke a scan command first. Then the HoloLens takes a photo which is then processed by a cloud vision service to find objects in the photo and thus in the real world. Found objects are highlighted in augmented reality as targets and can be approached with the UAV. For this, users only have to command the UAV to fly to the object they want, Therefore, users only need to scan and select targets for the UAV; most other work is automated. The UAV control was evaluated in a study with 26 participants with regard to user satisfaction as well as mental and physical demand. The results of the evaluation showed, that users would use such a UAV control in a smart home environment if they could benefit from it.

In future work, it would be interesting to combine the strengths of the HoloLens cameras with the versatility of the UAV such that the UAV can act even more autonomously. Then, it would be possible to transfer the scan function to the UAV and users would not even need to walk around. Additionally, other functionalities like displaying and changing paths as well as calculating more complicated paths, including collision detection and avoidance are necessary for smart home integration.

In general, the UAV not only needs to be enhanced with functionalities enhancing versatility and providing use-cases in a smart home environment but more research on human-UAV interaction is also required to pave the way for UAVs integrated into smart homes in the future. Especially aspects such as safety, privacy, and security, will play an important role in the interaction between humans and UAVs.

However, the work-wise of the prototype can be reused as a foundation for future smart-home-integrated UAVs since it contains some essential features like semi-automatic target detection of real-world objects and a way to command UAVs to precise positions with voice commands.

REFERENCES

[1] Zahra Ashktorab, Mohit Jain, Q. Vera Liao, and Justin D. Weisz. 2019. Resilient Chatbots: Repair Strategy Preferences for Conversational Breakdowns. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300484

[2] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.

[3] Louis-Pierre Bergé, Nabil Aouf, Thierry Duval, and Gilles Coppin. 2016. Generation and VR visualization of 3D point clouds for drone target validation assisted by an operator. In *2016 8th Computer Science and Electronic Engineering Conference, CEEC 2016, Colchester, UK, September 28-30, 2016*. IEEE, 66–70. https://doi.org/10.1109/CEEC.2016.7835890

[4] Kamel Boudjit and C Larbes. 2016. Detection and target tracking with a quadrotor using fuzzy logic. In *2016 8th International Conference on Modelling, Identification and Control (ICMIC)*. IEEE, 127–132.

[5] Endri Bregu, Nicola Casamassima, Daniel Cantoni, Luca Mottola, and Kamin Whitehouse. 2016. Reactive control of autonomous drones. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*. 207–219.

[6] Widodo Budiharto, Alexander AS Gunawan, Jarot S Suroso, Andry Chowanda, Aurello Patrik, and Gaudi Utama. 2018. Fast object detection for quadcopter drone using deep learning. In *2018 3rd international conference on computer and communication systems (ICCCS)*. IEEE, 192–195.

[7] Chien-Fang Chen, Kang-Ping Liu, and Neng-Hao Yu. 2015. Exploring interaction modalities for a selfie drone. In *SIGGRAPH Asia 2015 Posters*. 1–2.

[8] Martin Cooney, Francesco Zanlungo, Shuichi Nishio, and Hiroshi Ishiguro. 2012. Designing a flying humanoid robot (FHR): effects of flight on interactive communication. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 364–371.

[9] Liyanage C De Silva. 2008. Audiovisual sensing of human movements for home-care and security in a smart environment. *International Journal on Smart Sensing and Intelligent Systems* 1, 1 (2008), 220–245.

[10] Liyanage C De Silva, Chamin Morikawa, and Iskandar M Petra. 2012. State of the art of smart homes. *Engineering Applications of Artificial Intelligence* 25, 7 (2012), 1313–1321.

[11] Stephan Diederich, Alfred Benedikt Brendel, Stefan Morana, and Lutz Kolbe. 2022. On the design of and interaction with conversational agents: An organizing and assessing review of human-computer interaction research. *Journal of the Association for Information Systems* 23, 1 (2022), 96–138.

[12] Okan Erat, Werner Alexander Isop, Denis Kalkofen, and Dieter Schmalstieg. 2018. Drone-Augmented Human Vision: Exocentric Control for Drones Exploring Hidden Areas. *IEEE Trans. Vis. Comput. Graph.* 24, 4 (2018), 1437–1446. https://doi.org/10.1109/TVCG.2018.2794058

[13] Asbjørn Følstad and Petter Bae Brandtzæg. 2017. Chatbots and the new world of HCI. *interactions* 24, 4 (2017), 38–42.

[14] Wojciech Giernacki, Mateusz Skwierczynski, Wojciech Witwicki, Pawel Wronski, and Piotr Kozierski. 2017. Crazyflie 2.0 quadrotor as a platform for research and education in robotics and control engineering. In *22nd International Conference on Methods and Models in Automation and Robotics, MMAR 2017, Międzyzdroje, Poland, August 28-31, 2017*. IEEE, 37–42. https://doi.org/10.1109/MMAR.2017.8046794

[15] Vaishnavi S Gunge and Pratibha S Yalagi. 2016. Smart home automation: a literature review. *International Journal of Computer Applications* 975, 8887-8891 (2016).

[16] Richard Harper. 2006. *Inside the smart home*. Springer Science & Business Media.

[17] Viviane Herdel, Lee J. Yamin, and Jessica R. Cauchard. 2022. Above and Beyond: A Scoping Review of Domains and Applications for Human-Drone Interaction. In *CHI '22: CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 29 April 2022 - 5 May 2022*, Simone D. J. Barbosa, Cliff Lampe, Caroline Appert, David A. Shamma, Steven Mark Drucker, Julie R. Williamson, and Koji Yatani (Eds.). ACM, 463:1–463:22. https://doi.org/10.1145/3491102.3501881

[18] Pieter Hintjens. "O'Reilly Media, Inc.", 2013. *ZeroMQ: messaging for many applications*. "O'Reilly Media, Inc.".

[19] Eija Honkavaara, Heikki Saari, Jere Kaivosoja, Ilkka Pölönen, Teemu Hakala, Paula Litkey, Jussi Mäkynen, and Liisa Pesonen. 2013. Processing and Assessment of Spectrometric, Stereoscopic Imagery Collected Using a Lightweight UAV Spectral Camera for Precision Agriculture. *Remote. Sens.* 5, 10 (2013), 5006–5039. https://doi.org/10.3390/rs5105006

[20] Baichuan Huang, Deniz Bayazit, Daniel Ullman, Nakul Gopalan, and Stefanie Tellex. 2019. Flight, Camera, Action! Using Natural Language and Mixed Reality to Control a Drone. In *International Conference on Robotics and Automation, ICRA 2019, Montreal, QC, Canada, May 20-24, 2019*. IEEE, 6949–6956. https://doi.org/10.1109/ICRA.2019.8794200

[21] Xiyue Jia, Yining Cao, David O'Connor, Jin Zhu, Daniel CW Tsang, Bin Zou, and Deyi Hou. Elsevier, 2021. Mapping soil pollution by using drone image recognition and machine learning at an arsenic-contaminated agricultural field. *Environmental Pollution* 270 (Elsevier, 2021), 116281.

[22] Brennan Jones, Kody R. Dillman, Richard Tang, Anthony Tang, Ehud Sharlin, Lora Oehlberg, Carman Neustaedter, and Scott Bateman. 2016. Elevating Communication, Collaboration, and Shared Experiences in Mobile Video through Drones. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems, DIS '16, Brisbane, QLD, Australia, June 04 - 08, 2016*, Marcus Foth, Wendy Ju, Ronald Schroeter, and Stephen Viller (Eds.). ACM, 1123–1135. https://doi.org/10.1145/2901790.2901847

[23] Sunggoo Jung, Sunyou Hwang, Heemin Shin, and David Hyunchul Shim. 2018. Perception, Guidance, and Navigation for Indoor Autonomous Drone Racing Using Deep Learning. *IEEE Robotics Autom. Lett.* 3, 3 (2018), 2539–2544. https://doi.org/10.1109/LRA.2018.2808368

[24] Md Nafiz Hasan Khan and Carman Neustaedter. 2019. An exploratory study of the use of drones for assisting firefighters during emergency situations. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–14.

[25] Konstantinos Konstantoudakis, Kyriaki Christaki, Dimitrios Tsiakmakis, Dimitrios Sainidis, Georgios Albanis, Anastasios Dimou, and Petros Daras. MDPI, 2022. Drone Control in AR: An Intuitive System for Single-Handed Gesture Control, Drone Tracking, and Contextualized Camera Feed Visualization in Augmented Reality. *Drones* 6, 2 (MDPI, 2022), 43.

[26] Przemyslaw Mariusz Kornatowski, Mir Feroskhan, William J. Stewart, and Dario Floreano. 2020. A Morphing Cargo Drone for Safe Flight in Proximity of Humans. *IEEE Robotics Autom. Lett.* 5, 3 (2020), 4233–4240. https://doi.org/10.1109/LRA.2020.2993757

[27] Przemyslaw M. Kornatowski, Stefano Mintchev, and Dario Floreano. 2017. An origami-inspired cargo drone. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2017, Vancouver, BC, Canada, September 24-28, 2017*. IEEE, 6855–6862. https://doi.org/10.1109/IROS.2017.8206607

[28] Sven Laumer, Fabian Gubler, Ana Racheva, and Christian Maier. 2019. Use Cases for Conversational Agents : An Interview-based Study. In *AMCIS 2019 Proceedings*. AIS Electronic Library (AISeL), 10. https://fis.uni-bamberg.de/handle/uniba/50690

[29] Eun Ju Lee, Sang Young Shin, Byoung Chul Ko, and Chunho Chang. Elsevier, 2016. Early sinkhole detection using a drone-based thermal camera and image processing. *Infrared Physics & Technology* 78 (Elsevier, 2016), 223–232.

[30] Sang-won Leigh, Harshit Agrawal, and Pattie Maes. 2016. A flying pantograph: Interleaving expressivity of human and machine. In *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*. 653–657.

[31] James Lester, Karl Branting, and Bradford Mott. 2004. Conversational agents. *The practical handbook of internet computing* (2004), 220–240.

[32] Marcello M Mariani, Novin Hashemi, and Jochen Wirtz. 2023. Artificial intelligence empowered conversational agents: A systematic literature review and research agenda. *Journal of Business Research* 161 (2023), 113838.

[33] Davit Marikyan, Savvas Papagiannidis, and Eleftherios Alamanos. 2019. A systematic review of the smart home literature: A user perspective. *Technological Forecasting and Social Change* 138 (2019), 139–154.

[34] Kimberly McGuire, Guido de Croon, Christophe De Wagter, Karl Tuyls, and Hilbert J. Kappen. 2017. Efficient Optical Flow and Stereo Vision for Velocity Estimation and Obstacle Avoidance on an Autonomous Pocket Drone. *IEEE Robotics Autom. Lett.* 2, 2 (2017), 1070–1076. https://doi.org/10.1109/LRA.2017.2658940

[35] Michael F McTear. 2002. Spoken dialogue technology: enabling the conversational user interface. *ACM Computing Surveys (CSUR)* 34, 1 (2002), 90–169.

[36] Susanna Paasovaara, Andrés Lucero, and Thomas Olsson. 2016. Outlining the design space of playful interactions between nearby strangers. In *Proceedings of the 20th International Academic Mindtrek Conference*. 216–225.

[37] Pan Peining, Gary Tan, and AA Phyo Wai. 2017. Evaluation of consumer-grade EEG headsets for BCI drone control. In *IRC Conference on Science, Engineering, and Technology*, Vol. 80.

[38] Atyanti Dyah Prabaswari, Chancard Basumerda, and Bagus Wahyu Utomo. 2019. The mental workload analysis of staff in study program of private educational organization. In *IOP Conference Series: Materials Science and Engineering*, Vol. 528. IOP Publishing, 012018.

[39] Philip Pratt, Matthew Ives, Graham Lawton, Jonathan Simmons, Nasko Radev, Liana Spyropoulou, and Dimitri Amiras. 2018. Through the HoloLens™ looking glass: augmented reality for extremity reconstruction surgery using 3D vascular models with perforating vessels. *European radiology experimental* 2 (2018), 1–7.

[40] Jacopo Primicerio, Salvatore Filippo Di Gennaro, Edoardo Fiorillo, Lorenzo Genesio, Emanuele Lugato, Alessandro Matese, and Francesco Primo Vaccari. Springer, 2012. A flexible unmanned aerial vehicle for precision agriculture. *Precision Agriculture* 13, 4 (Springer, 2012), 517–523.

[41] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I Hong. 2018. " Hey Alexa, What's Up?" A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 designing interactive systems conference*. 857–868.

[42] Biljana L Risteska Stojkoska and Kire V Trivodaliev. 2017. A review of Internet of Things for smart home: Challenges and solutions. *Journal of cleaner production* 140 (2017), 1454–1464.

[43] Andrzej Szajna, Roman Stryjski, Waldemar Wozniak, Norbert Chamier-Gliszczynski, and Mariusz Kostrzewski. 2020. Assessment of Augmented Reality in Manual Wiring Production Process with Use of Mobile AR Glasses. *Sensors* 20, 17 (2020), 4755. https://doi.org/10.3390/s20174755

[44] Dorin Ungureanu, Federica Bogo, Silvano Galliani, Pooja Sama, Xin Duan, Casey Meekhof, Jan Stühmer, Thomas J. Cashman, Bugra Tekin, Johannes L. Schönberger, Pawel Olszta, and Marc Pollefeys. 2020. HoloLens 2 Research Mode as a Tool for Computer Vision Research. *CoRR* abs/2008.11239 (2020). arXiv:2008.11239 https://arxiv.org/abs/2008.11239

[45] Kristen Vaccaro, Tanvi Agarwalla, Sunaya Shivakumar, and Ranjitha Kumar. 2018. Designing the Future of Personal Fashion *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3173574.3174201

[46] Alexander C. Woods and Hung M. La. 2015. Dynamic Target Tracking and Obstacle Avoidance using a Drone. In *Advances in Visual Computing - 11th International Symposium, ISVC 2015, Las Vegas, NV, USA, December 14-16, 2015, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 9474)*, George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Ioannis T. Pavlidis, Rogério Schmidt Feris, Tim McGraw, Mark Elendt, Regis Kopper, Eric D. Ragan, Zhao Ye, and Gunther H. Weber (Eds.). Springer, 857–866. https://doi.org/10.1007/978-3-319-27857-5_76

[47] Stephen Xia, Rishikanth Chandrasekaran, Yanchen Liu, Chenye Yang, Tajana Simunic Rosing, and Xiaofan Jiang. 2021. A Drone-based System for Intelligent and Autonomous Homes. In *SenSys '21: The 19th ACM Conference on Embedded Networked Sensor Systems, Coimbra, Portugal, November 15 - 17, 2021*, Jorge Sá Silva, Fernando Boavida, André Rodrigues, Andrew Markham, and Rong Zheng (Eds.). ACM, 349–350. https://doi.org/10.1145/3485730.3492881