Natural User Interfaces for Human-Drone Multi-Modal Interaction

Ramón A. Suárez Fernández[1], Jose Luis Sanchez-Lopez[1], Carlos Sampedro[1],
Hriday Bavle[1], Martin Molina[2], and Pascual Campoy[1]

***Abstract*— Personal drones are becoming part of every day life. To fully integrate them in society, it is crucial to design safe and intuitive ways to interact with these aerial systems. The recent advances on Natural User Interfaces (NUIs) intend to make use of human innate features, such as speech, gestures and vision to interact with technology the way humans would interact with each other.**

In this paper, several NUI strategies are developed and implemented along with computer vision techniques, for intuitive and natural human-quadrotor interaction in indoor GPS-denied environments. These strategies include speech, body position, hand gesture and visual marker interactions used to directly command tasks to the drone. The NUIs presented are based on devices like the Leap Motion Controller, microphones and small size monocular on-board cameras which are unnoticeable to the user. Thanks to this, the users can choose the most intuitive and effective type of interaction for their application. Additionally, the strategies proposed allow for multi-modal interaction between multiple users and the drone by being able to integrate several of these interfaces in one single application as is shown in various real flight experiments performed with non-expert users.

I. INTRODUCTION

In recent years, the demand for drones in civilian and non-civilian applications has caused the commercial drone markets to grow exponentially. Newly created businesses and technology solutions appear every day with the hope of competing in this field. Operators have an important role in the control schemes of drone platforms given that these fly are semi-autonomous. Aerial vehicles play a fundamental role in several applications spanning from surveillance/search-and-rescue all the way to entertainment [35]. Conventional platforms require direct physical contact for communication with the interfacing systems keyboard or mouse, limiting the scope and dimension of interaction drastically. With the advent of innovative gesture-based NUIs a new frontier of communication techniques have evolved. This will soon allow non-expert users, who have little knowledge about keyboard and the system, to interact and operate the robots using natural gestures [12]. Thus, creating a new communicating language and offering exciting new applications for *social* drones that blend in and share participation in several tasks.

*The authors would like to thank the Consejo Superior de Investigaciones Cientificas (CSIC) of Spain for the JAE-Predoctoral scholarships of one of the authors and their research stays, and the Spanish Ministry of Science MICYT DPI2014-60139-R for project funding, as well as the Spanish Ministry for Education, Culture and Sports for funding the international research stay of one of the authors".

[1]Computer Vision Group, Centre for Automation and Robotics, CSIC-UPM (Spain). {jl.sanchez, pascual.campoy}@upm.es. www.vision4uav.eu

[2]Department of Artificial Intelligence, Technical University of Madrid (UPM) (Spain)

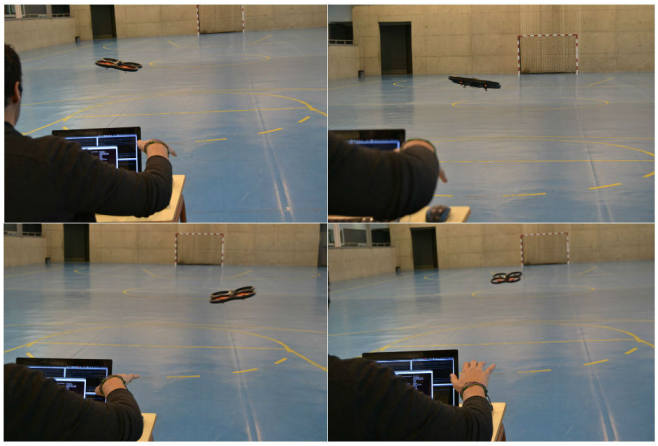


Fig. 1: Human-drone hand gesture interaction being tested in indoor environments using the Leap Motion sensor.

Natural user interfaces have been researched since the 1980s, for instance in [4], the author used gestures and voice commands for control of a graphical user interface (GUI). The two main mediums to implement reliable human-drone interaction (HDI) are Voice and gesture based NUIs. Since the latter represents direct expression of mental concepts, it is the most preferred in literature [25].

The wide range of hand/arm and body gestures that can be recognized, as well as the verbal vocabulary that can be communicated to the drone, can offer unique opportunities for developing new and captivating types of HDIs [21]. Hence, new gesture-based NUIs have been increasingly implemented in several scenarios encompassing, for example, interaction in virtual reality environments [13], controlling robotic fish [27], interacting with robots [14], etc. Recent technological advances in the areas of controllers and sensors used as input devices for HDI, mixed together with the relatively low cost of drones equipped with on-board cameras can be exploited as affordable devices that support the design and implementation of new kinds of NUIs.

This work aims at developing NUIs for performing efficient and natural HDI and control. These interfaces allow autonomous drone navigation in GPS-denied environments by using the users body position, hand gestures, visual markers and/or speech. Various sensors are used as input devices, to recognize users high level commands which are then used to control the drone platform, these include on-board cameras for body position and visual marker commands, the Leap Motion hand tracking controller for hand gesture control,

and audio input/output for speech recognition and control.

There are other works found in the literature, aimed to explore this type of interaction. In [19] a drone is teleoperated by sending discrete control commands given by static arm gesture recognition techniques. Whereas, the flying machine arena in ETH Zürich proposed a solution to directly map the users arm coordinates to the drone's position using the Microsoft Kinect™ [2]. And in [10] and [22] the authors explore the idea of using a runners body to exercise with a drone companion.

Nevertheless, the approach taken in this paper differs from the previous examples in the manner in which user commands are sent to the drone and interact before and during flight. While many solutions have used depth cameras like the Kinect and similar devices, this work introduces sensors like the Leap Motion Controller which is explicitly targeted for hand gesture recognition. This sensor directly computes the position of fingertips and hand orientation used for control, giving a more intuitive flight experience to the user with the palm of their hand. On the other hand, by taking advantage of the drones' cheap on-board front facing cameras, body position estimation and tracking can give the user the ability to interact on a personal level where the drone follows the user and moves according to the users position. This kind of interaction is especially interesting for cinematography applications using front facing camera views.

Furthermore, this paper introduces interactions via visual markers and speech. Using visual markers and/or speech to send commands to the drone, allows the user to interact with the system either from a landed state or mid-flight and perform tasks such as *take-off, move, flip, hover, land*, etc. This multi-modal interaction gives a higher degree of cooperation where the user can communicate seamlessly between modalities permitting stimulating and safe user experiences for non-skilled users.

The remainder of the paper is organized as follows. In Section II, background information is presented. This includes a brief history on NUIs as opposed to GUIs and quadrotor platforms. Section III will explain the system architecture for the graphical and natural user interfaces implemented. Finally, sections IV and V will detail the real flight tests performed and conclusions respectively.

II. BACKGROUND

A. NUI vs. GUI

In HDI frameworks, above all in demanding operations such as search-and-rescue or indoor navigation, it is crucial for humans to be able to interact and command robots in natural and efficient ways. To allow effective operation and control of the machine from the human end, interactions between the humans and machines must occur. This interaction takes place in the user interface (UI).

The first UIs were simple designs, not intuitive for the users. With newly developed hardware becoming ever so common, new technologies were being developed which meant researchers had to create new forms of interacting

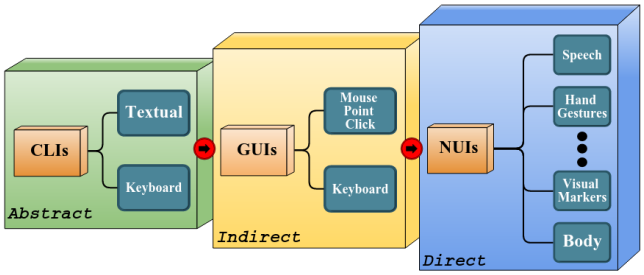


Fig. 2: The User Interface evolution. Command line interfaces began the UI revolution followed by a more indirect GUI. The most recent user interface is the NUI.

with machines. Hence a growing interest in designing new types of interfaces has developed over the decades. As stated in [37], these types can initially be divided into command line interface (CLI) succeeded by the GUI (Fig. 2). Nevertheless, the fact that in CLIs operators have to interact with the systems by typing preprogrammed keywords into a command prompt can lead to novice users feeling completely overwrought when experiencing these interfaces for the first time [24].

The GUI, that is still used today, produces an indirect but expected mode of interaction by using what is commonly referred to as WIMP (Windows, Icons, Menus, Pointer) [37] [36], a set of user interface elements that serve as user inputs and machine outputs. The WIMP style of interaction, as coined by Merzouga Wilberts in 1980, gives a stable and universal face to computing where simple commands can be chained together to set about a group of commands that would have taken several command line instructions to complete while automatically providing direct feedback of what users do. In comparison to CLIs, these interfaces represent a lower obstacle for users since recognizing and choosing commands is easier than remembering and typing [37]. The properties of WIMP GUIs provide users a clearer idea of what actions and processes are available in the system as well as what their effects might be, this allows users to have a sense of achievement about their interactions with computer programs [36]. Despite the fact that GUIs have been very prosperous and controlled both human-machine interface (HMI) research and the marketplace for most of three decades [20], these interfaces are not the most efficient option, given that there is still a barrier between the communication from human to machine. Consequently, there is an ever growing demand to create more immerse UIs that take full advantage of modern technologies allowing users to be able to feel fully integrated into the devices they use [32]. One step towards achieving this is the Natural User Interface.

[30] describes NUIs as types of interfaces that allow users to engage with machines in a similar way they would interact with the real world through using body movements, hands or even voice. Unlike GUIs where users had to use keys, buttons or a computer mouse, now language, touch, or body movements are used to control a device. Thus, the user can

interact directly with the elements on the screen without an additional device, such as the mouse. Thereby a seamless interaction between humans and computers has been created by which you can handle virtual or real objects in a realistic manner. Most NUIs rely on additional equipment for suitable and efficient interaction, however, these devices tend to be so unnoticeable that they appear invisible to the user.

B. Quadrotors and the ARDrone 2.0 Platform

Quadrotor platforms have four independently fixed pitch controlled rotors that allow the operator to control the vehicle in height, orientation and translation. Due to their basic mechanical designs, low price and assortment of sizes, these platforms are suitable to be interacted with in safe, controlled and fun environments. This type of unmanned aerial system (UAS) is a useful tool for researchers to test and validate any number of concepts in fields such as control, state estimation, sensor fusion, and mobile robotics.

Recently, research centers and universities have shown the potential of the quadrotor platforms by performing more and more complex aerial maneuvers [18][23]. Thanks to these attributes these platforms are suitable for tasks such as surveillance, search-and-rescue and inspection. Among the commercial platforms available (favored by the research institutions) some frequently used are the Ascending Technologies[1] Pelican and the Parrot[2] AR.Drone 2.0.

The quadrotor used in this work was the Parrot AR.Drone 2.0 a small low cost platform that is commercially available for around 300. This quadrotor is widely used in research groups for features such as on-board cameras and stable hovering. A more complete description can be found in [28]. To carry out the autonomous operation of quadrotors while researching on HDI (GUIs and NUIs), Aerostack[3][4] was employed. A full description of this architecture and software framework is out of the scope of the paper and can be found in [33] and [34].

Aerostack (see Fig. 3) has a set of available actions (like *take off*, *land*, *hover*, *move*, *turn in yaw*, etc.) and behaviors (such as *recognize object*, *recognize markers*) that together with its performance and state, can be directly used ensuring its fully autonomous operation. Additionally, Aerostack provides the value of all its measured and estimated magnitudes, as well as all the internal states or commands of every component, that can be directly used by the user.

III. Researched Human Drone Interactions

In this section the implemented user interfaces will be explained. As mentioned in the previous section, GUIs have had a big impact on HMI and are a great tool for robotics, thus the need to implement a robust and reliable interface for the drone platforms which will be explained in Section III-A. After the interaction with the drones evolved, the need for more natural interfaces grew. These are described in Section III-B.

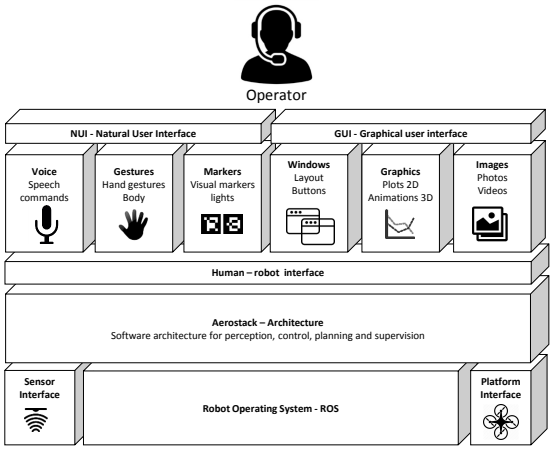


Fig. 3: The NUIs were developed using the Aerostack framework, these interfaces communicate via the HRI layer with the Aerostack software architecture. This software, available online, then sends desired commands to the platform.

A. Graphical User Interface (GUI)

This section describes how the operator can interact with quadrotor platforms using the developed GUI. In general, a GUI provides some functions to help operators in certain tasks that are difficult to be supported by a NUI. For example, they correspond to tasks when the operator requires detailed information such as vehicle set up or mission monitoring at software level (e.g., during software maintenance).

The GUI allows the interaction with the vehicle, observing the states and dynamics and presents graphical views and images to help the user to understand both the external and internal behavior of the vehicle.

In general, the operator can use a GUI to perform the following types of tasks:

- *Specify drone behavior in advance (vehicle set up)*
- *Monitor drone behavior during a mission*
- *Operate manually with simple movements*
- *Collect data for later use*

Fig. 4 shows a sample screen of the user interface with the windows layout with main parts: the control panel on the upper left side of the image, the dynamics viewer on the lower left side, the windows for detailed content on the right hand side of the screen with different tabs (parameters, the camera viewer or the performance monitor). At the top, there are drop-down menus (file, view, settings, etc.) to perform additional tasks.

The following sections describe the different parts of the graphical user interface.

• *The control panel:* The control panel shows the general state of the system and the main control commands. This panel provides a summary of information considered critical that needs to be permanently present on the screen.

[1]Online:`http://wiki.asctec.de`
[2]Online:`http://ardrone2.parrot.com/`
[3]Aerostack webpage: `http://www.aerostack.org/`
[4]GitHub :`https://github.com/Vision4UAV/Aerostack`

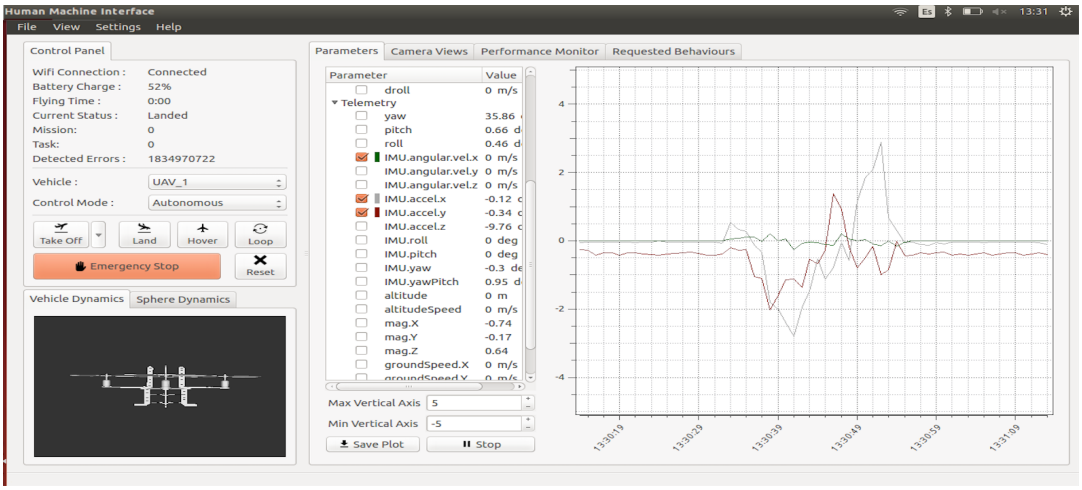
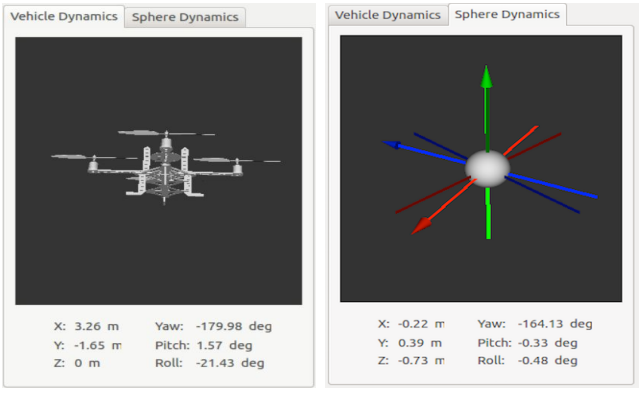


Fig. 4: Sample screen of the graphical user interface.

• *Parameter viewer:* In order to monitor the vehicle's behavior in detail during a mission, the operator can observe the values of numerical parameters and display multiple plots of the parameters selected in real time to analyze and compare their values.

• *Camera viewer:* The camera viewer shows pictures and/or video images captured by the aerial vehicle during flight.

• *Requested behaviors viewer:* The operator can use the behavior viewer to consult, and request the activation of specific behaviors. The GUI shows a list of available behaviors for the vehicle and indicates for each one its state.



(a) The Vehicle Dynamics. (b) The sphere Dynamics.

Fig. 5: The Dynamics Viewer.

• *The dynamics viewer:* The GUI includes a viewer that shows in animated 3D views the dynamics of the vehicle. This can be used by the operator, for example, to better guide the vehicle during manual operations of the vehicle with the keyboard.

This viewer presents two animated images as seen in Fig. 5: (1) the vehicle dynamics, a 3D representation of the vehicle (Fig. 5(a)) and (2) the sphere dynamics, a sphere with orientation axis (Fig. 5(b)). In the sphere representation, there are three fixed axis that represent the reference system, and three variable axis that represent the orientation changes.

B. Natural User Interfaces (NUIs)

On the other hand, there are operator tasks where a NUI can provide a more efficient communication compared to a conventional GUI. Natural communication can include using gestures to guide the vehicle during manual operation which are easier to learn or voice commands that can be used more efficiently in combination with other communication modes.

In general, a communication based on NUI can be especially useful in human-drone cooperative work for certain complex missions in dynamic environments where the partial information used by the vehicle can be complemented with human decisions.

In the following sections a description of the types of NUIs implemented and the setup used for each interaction will be given. First, the vision based NUIs will be explained in Section III-B.1 and Section III-B.2 starting with *body* and followed by *marker* interaction. Afterwards, *hand* interaction is described in Section III-B.3 and finally *speech* interaction is detailed in Section III-B.4.

1) Visual Body Interaction: The drone, equipped with an on-board camera, uses computer vision algorithms to detect a person and track it in the image plane. With the previous knowledge of the approximate dimensions of the object tracked, the drone is able to reconstruct the 3D relative position of the object with respect to the drone. A control algorithm sends commands to the drone ensuring it maintains the distance (x, y and z) and point of view (yaw angle). This controller was already developed by the authors in [26] where more details can be consulted.

Fig. 6 presents the high-level description of the proposed NUI. In the context of the presented work, this controller is going to be used to explore a new kind of human-drone interaction. With this purpose, the computer vision algorithm is enforced to detect and track the person with whom it is interacting. The drone then follows the person, keeping the distance and point of view, even if the person moves or runs.

Fig. 6: Description of High-level setup for the proposed body position NUI.

For safety reasons, the quadrotor is always employed with the indoors hull to protect the person against the propellers.

This NUI demonstrates one of the most primitive behaviors of animals. Baby animals, by instinct, follow their parents everywhere, this interaction is similar to how the presented drone is behaving. The author's belief on this instinctive NUI, led them to test it over a great sample of different ages (from children to elderly).

2) Visual Marker Interaction: Visual cues of color, depth and motion are in many species, specially in humans, a large source of information in how the world is perceived [29]. Using visual cues or markers is not an uncommon practice in robotics since these take use of arguably the most important sensor in drones or robotics, the camera. Simple monochromatic cameras can be used for accurate and reliable target tracking or detection. For example, in [6], the authors use tags to control an underwater robot navigating in a pool.

In this type of interaction between the human and the drone, the user manipulates markers (otherwise regarded as tags in the literature) to *show* the drone what to do. Since no additional device apart from the on-board front facing camera is needed, the interaction is effortless and the user feels integrated into the decision making process of commanding the drone. By using precisely engineered markers, robust and accurate interaction is obtained while maintaining a large level of convenience for the user.

Fig. 7 shows a representation of the real system in use. The idea behind this is that a non-expert user can pick up a predefined set of visual command markers and interact in a safe and entertaining way. The visual markers employed serve as fiducial markers in the video environment. These are visual patterns that allow robust and accurate detection and whose pose can be precisely estimated [6]. These markers rely on a specific pixel pattern that uniquely encodes information which is needed for the detection algorithms to operate.

There are many available fiducials in literature. In [15], the authors develop the ReacTIVision amoeba marker which is based on blob detection optimized by genetic algorithms. The ARTag system [8] is based in the idea of binary code and implements edge-based square detection methods.

In this work the visual markers used are the ArUco Tags developed by the AVA group in the University of Cordoba [9]. The tag system proposed is also based on binary coding. Nevertheless, instead of using a predefined set of markers, they developed a way of producing configurable marker dictionaries (with arbitrary size and number of markers). This allows to create different markers of varying sizes, thus expanding the command list that can be sent to the drone, making this an evolving communication system.

The command process is as follows: when the markers appear in the on-board cameras filed of view, they are detected and identified using the vision processing algorithms in the ground station. Depending on the identifiers of the detected markers, different tasks are sent via the Aerostack for the drone to perform actions such as *take off*, *hover*, *land*, etc.

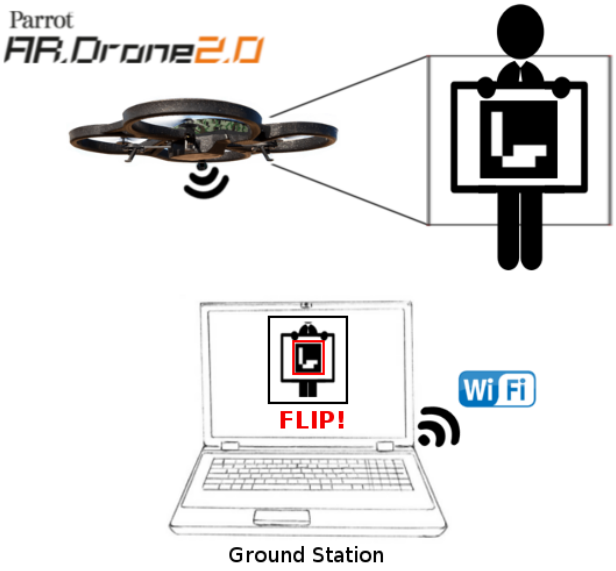


Fig. 7: High-Leved Description of Visual Marker NUI

3) Hand Gesture Interaction: Hand gestures are frequently considered the most expressive and in so, the most often used in the literature for new NUIs. These gestures involve: 1- a posture: normally expressed by a lack of hand movement with predetermined finger configurations and 2- a gesture: where hand motion is dynamic [16].

The interpreting of gestures by the NUI requires that the configurations (static and/or dynamic) of the human hand and even arm, be measurable by some device. Initial attempts at hand/arm based NUIs were known as *glove-based devices*

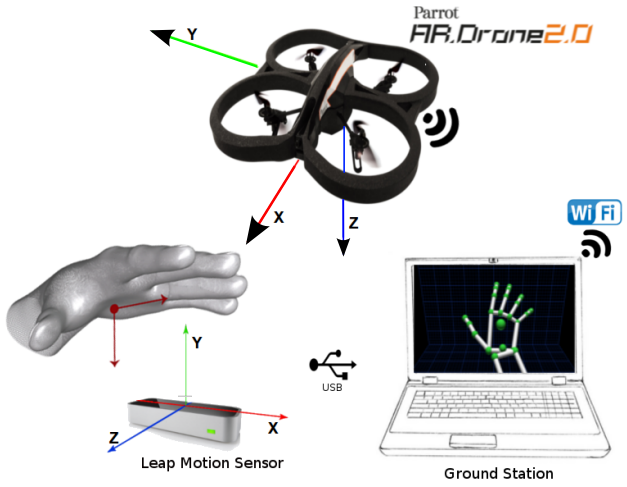


Fig. 8: High-Level Description of Hand Gesture NUI

[3][7]. These works depended on unwieldy sensors that directly measured spatial position and joint angles. Users found these devices to impede the interaction between the user and the computer controlled environment [16]. The need for more natural interaction between the human and the machine has spawned research into the use of other devices.

Given the many software libraries available and the relative low-cost of the devices, most research into hand gesture based NUIs has drifted into using sensors equipped with depth cameras such as the Kinect [35] [27] [19]. For example, In [5], the authors develop a gesture recognition system based on depth imagery and create a depth-based hand gesture database for control of drones. [1] successfully tested in real world scenarios a tour-guide robot that recognized and provided feedback for user hand gesture commands or augmented reality virtual button selection.

However, such devices present problems when it comes to accurate recognition of depth-based hand gestures which include, reduced resolution, high noise and missing data. These deficiencies makes it infeasible to extract reliable data of accurate hand/finger poses [5]. Within the new design tools that interpret hand movement and poses in a three dimensional space is the Leap Motion.

The Leap Motion controller is an 8 cm long USB connected sensor designed to track hand and finger motions in a small working space. The device is intended for consumer use and requires minimal setup on the host computer. This provides accessible means of controlling practically any interface with the palm of your hand.

The sensor works by projecting infrared light upward and detecting reflections using monochromatic infrared cameras. It's field-of-view (FOV) extends from 25mm to 600mm with a 150°spread from the device, with a frame-rate of roughly 200fps and a precision of 1/100mm per finger [11].

The proposed hand gesture NUI description can be seen in Fig. 8. The Leap Motion computes the orientation and position of the users hand relative to it's own axes. Thus, with the configuration shown in Fig. 8, the pitch is the angle between the negative z-axis and the projection of the vector onto the y-z plane. In other words, pitch, roll and yaw represent rotations around the x, z and y axes respectively. In contrast, the drones axes follow a North-East-Down (NED) configuration, which means pitch, roll and yaw are rotations around the y, x and z axes respectively.

One way of using this information to interact with the drone is to send pitch, roll, yaw rate and thrust commands to the drone by directly mapping the orientations and position of the hand to the drones coordinate system. Another possibility is to send higher level commands such as velocity or position. After conducting several experiments, the results show that a direct transformation between the movement of the hand and the movement of the drone felt more instinctive for users.

Fig. 9: High-Level Description of Speech Command NUI

4) Speech Command Interaction: This type of interaction has become more and more common for HCIs due to the growing number of speech recognition softwares and toolkits that are readily available. Early examples of speech recognition interfaces were mainly used for speech-to-text applications. Products such as Dragon Naturally Speaking allow the user to dictate and have speech transcribed as written text, have a document synthesized as an audio stream, or issue commands that are recognized as such by the program. These type of applications bridge the gap between the spoken word and it's written form and offer hands/eyes-free interfaces that are intuitive and appealing to the user. Hence, this type of interaction, when used in drone applications, can enhance the user's experience by being able to devote all their visual attention at commanding tasks to the drone.

Few examples of speech interaction with drones can be found in literature. In [31] a voice controller was developed that could recognize commands sent to a fixed wing semi-autonomous UAS using a PDA. Real-flight tests with their interface showed that ambient wind noise and conversation

can lower the reliability of the voice recognition system. In [13], Jones *et al.* conducted an exploratory study of gesture and speech interfaces for interaction with robots in a simulated environment, which concluded that the test subjects generally preferred using lower-level commands such as *left* or *right* to command the drone. The development of an effective speech command interaction requires prior in-depth knowledge of the tasks that can be performed and who the end-user will be. These interfaces need to respond to input reliably or they may be rejected by the user.

Fig. 9 shows the high-level setup of the proposed speech command NUI. As can be seen, the setup is simple. No devices other than a microphone are needed to command tasks to the drone. Voice commands are sent to the a Ground Station to be processed and later sent as tasks to the drone using the Aerostack.

Voice processing is done using the ROS package implementation of the Pocketsphinx library. The CMU Pocket Sphinx speech recognizer is the general term to describe a group of speech recognition systems based on hidden Markov models (HMM's) developed at Carnegie Mellon University [17]. This package automatically splits the incoming audio into utterances to be recognized. Currently, the recognizer requires a language model and dictionary file that can be automatically built from a corpus of sentences using the Online Sphinx Knowledge Base Tool[5]. This software allows to easily integrate new voice commands in order to expand the tasks or behaviors required for the drone.

A grammar of approximately fifteen commands has been developed that include, but are not limited to, *move forward*, *move backward*, *rotate right*, and the like. The software package listens for these simple one to three word tasks, and when a positive detection is made, a voice synthesizer was implemented to offer acknowledgement of the action in the present progressive tense: *taking off*, *moving forward*, etc.

IV. EXPERIMENTS AND RESULTS

In order to test and evaluate the performance of the aforementioned NUIs, a series of real-time experiments were conducted. These flight-tests consisted of individually testing the proposed interaction methods in controlled indoor environments without the aid of Motion Capture Systems, thus evaluating whether or not these interfaces fit the needs and expectations of seamless natural interaction between the user and the drone.

Apart from individually testing each interface method, a multi-modal interaction scenario is also tested by combining the interfaces and permitting the user decide which interaction to use at any given time.

A. Visual Body Interaction

In these experiments the authors encouraged users to interact and command the drone by changing the position of their body. Since the computer vision algorithms needed to automatically detect a person and track them is beyond the scope of this paper, the person to track is defined using a bounding box in a single frame. Afterwards, the algorithm tracks the person, learns their appearance and detects them when they appear in the frame. Thus, tracking is improved over time.

Fig. 10: This figure shows several users in different indoor scenarios interacting with the drone using the position of their body.

As can be seen in Fig. 10, users interacted with the drone in a number of ways including jumping in the air, turning around and even doing push-ups to test the limits of interaction. The feedback received from the users, specially those who had no previous interaction with aerial systems, was in general a positive one. Naturally, many users who tested this interaction felt doubtful at first, understandably thinking that this machine who was following them, could lose control at any moment. Nevertheless, after some time interacting with the drone, users acknowledged that having such a close one-to-one interaction/relationship with the drone gradually made them more and more comfortable being close to it. After the flights had ended, many people who had experienced this interaction ended up sharing interesting applications for this type of interface ranging from personal trainers to personal cameramen for news reporting and documentaries.

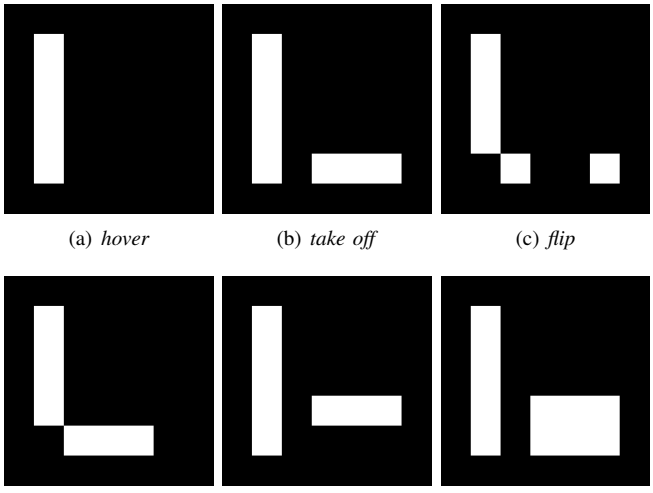
The results and the feedback was always the same: at the beginning, people were cautious and a little frightened. They remained static, just looking at the drone. Once the authors encouraged them to move away, they started with small quasi-static motions but quickly began to feel very confident and forgetting that it was a robot, threating it as a living being. Reports like *I felt like it was my friend* or *I want a pet like this* were very common and by the end people felt very upset when the drone ran out of battery.

B. Visual Marker Interaction

This interaction was performed by using hand-held visual markers developed with the ArUco tags. As was explained in Section III-B.2, each marker is unique. These markers

[5]Online: `http://www.speech.cs.cmu.edu/tools/lmtool-new.html`

are assigned a task to perform when the user presents it to the quadrotor. Fig. 11 shows some examples of the markers used in flight for actions such as *take off* (Fig. 11(b)) or *flip* (Fig. 11(c)).



(a) *hover* (b) *take off* (c) *flip*

(d) *follow me* (e) *land* (f) *follow trajectory*

Fig. 11: Examples of Visual Markers used in flight.

Fig. 12 shows users interacting with the drone using the visual marker NUI. For the test, users were given markers and were instructed they could use these in any way to interact with the aerial system. Since these markers included the flip action, this was the users preferred command to send to the drone. This test resulted in the interaction being described as fun and reliable. This may be due to the fact that the marker system employed is incredibly robust and users had instant response when commanding the quadrotor as opposed to body position and hand gesture interaction that the system has a slight delay in the response since it sends direct commands to the drone and not high-level tasks to perform.

This method of interaction extends the capabilities of the drone to perform robustly in applications where verbal of physical interaction is virtually non-existent. Among such applications is aiding individuals with disabilities. People with disabilities that are non-verbal or don't normally express their needs via verbal interactions need an alternative way to do so. Tools such as picture cards have been used to provide a way for these people to express themselves which, in some cases, can serve as a bridge to verbal communication. This principle of communication for people who aren't able to do it otherwise, can be employed in the use of drones, for example, to help in every day tasks or as a guide.

C. Hand Gesture Interaction

As was mentioned in Section III-B.3, this interaction is achieved with the use of the Leap Motion controller. Fig. 13 shows the gesture commands used for direct flight control.

User interaction with this method was not easy at first. Users had to get the feeling of the relationship between the

Fig. 12: Here the visual marker interaction is taking place. As can be seen in the bottom left image, the visual algorithm detects the arucos Ids, and perform the commanded tasks.

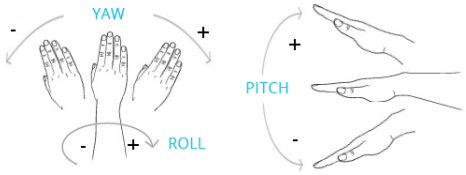


Fig. 13: Hand gesture commands used for drone flight. In this setup positive pitch, roll and yaw rate commands move the quadrotor backwards, right and clockwise respectively.

movement of their hand and the movement of the drone while having to compensate the different dynamics of the hand motions relative to the drones'. Nevertheless, experiencing the connection of the hand with the drone made this the most natural and fun method to interact amongst users. This response from the users may be due to the fact that nowadays, basic mundane devices such as smartphones or tablets employ this type of behavior in their UIs and people are accustomed to dealing with such interactions.

Fig. 14: Hand gesture interaction test flights

Users of different ages have used the proposed NUI as seen in Fig. 14 and Fig. 1. Most of the people who used the interface regarded the interaction as a game which felt intuitive and entertaining. This NUI can be used in many

different applications spanning from teleoperation of aerial systems in dangerous environments to quadrotor gaming systems.

D. Speech Command Interaction

This experiment was conducted to test the reliability and usefulness of the speech NUI for drone interaction. Table I details the commands that are presently recognized by the speech NUI. As was mentioned before, these commands are not fixed, they can be extended for any additional need the user might have in the future. The speech vocabulary was created to show the potential uses of the drone as well as to test the response of the system in real-flight scenarios.

Speech NUI Command List		
take off	*land*	*sleep*
hover	*move right*	*move left*
rotate right	*rotate left*	*arm*
move forward	*move backward*	*flip*
start visual body	*start visual marker*	*start speeh*
start hand gesture	*start trajectory*	*pause speech*

TABLE I: Speech NUI list of high-level commands.

In general users were drawn to this type of interaction. Most users expressed that talking to the drone felt like how one would talk to a pet. Or more precisely, how one felt while training a pet. These responses to the interaction were exalted by the fact that voice feedback was implemented using a voice synthesizer to reply when commands were correctly received.

E. Multi-Modal Interaction

In order to validate a full Natural User Interface system, multi-modal flight tests were performed using all of the NUIs proposed in this work. In this way, multiple users can interact with the drone in order to perform different tasks. In the test performed, one user was in charge of the speech interface and hand gesture interactions, while a non-skilled user was in charge of the body position interaction and the visual markers.

This scenario could easily represent rescue teams in disaster scenarios where a base operator stays with the ground station, monitoring the environment situation while interacting by voice or hand movements with the drone. The base operator would then guide and/or aid the rescue operators perform their mission. After some event has occurred, the rescue operators can signal the drone via visual markers to perform different tasks such as return home guiding the victims without having to exit the area and help others. A video demonstration of multiple users can be found in the following link: `https://youtu.be/-xLToLVE9qk`.

F. NUI Diffusion

These NUIs has been introduced in many shows and exhibitions for both general and technical public, including the 2014 and 2015 European Week of the Robotics, a 2015 Madrid TEDx entitled "Aerial Invasion?" and the 2015 European Researchers Night in Madrid[6]. Among the 400 members from the audience in the Researchers Night, few were allowed to engage with the drones without any prior knowledge of the abilities or commands that could be sent. Public acceptance was general, especially in the case of a handicapped teenager who was in a wheelchair, but, was able to show the markers to the drone and thus flew his first quadrotor platform.

This had not been the first time that these NUIs were used for the physically disabled to interact with drones. Previously, the proposed body position interface was used with disabled children from a special needs school in Madrid to teach them about robotics and specifically quadrotors in the 2014 Madrid Science Week[7].

As can be seen, these interfaces have been used in a variety of situations and tested with a number of different users in real flights which shows the potential and the reliability of these proposed NUIs.

V. CONCLUSIONS AND FUTURE WORK

Designing efficient, reliable and intuitive Natural User Interfaces is a key task in user centered design (UCD) philosophy. In this paper, several NUI strategies were proposed, whereby the interaction with drones evolved from touch to touch-less, by using *speech*, *hand gestures*, *body position* or *visual markers*. Adopting more affordable sensors, like the leap motion and small on-board monocular cameras, the general acceptance of the presented methods was demonstrated by users who validated them in real flights providing feedback of overall usability. These interfaces have also been tested in public events such as the 2015 European Researchers night and the 2014 and 2015 editions of Madrid Science Week where the overall comments from the users were positive and convenient since new and original applications for these aerial systems were being contributed.

The main strength of the proposed interaction methods is the ability to perform multi-modal interactions. With this, the user can employ any of these types of NUIs interchangeably to fulfill their application requirements, thus, expanding the applications in which new modes of HDI can be adopted and employed. Links to videos of real flight tests have been provided as well as information on the software architecture on which it was developed Aerostack. This work concludes that the use of NUIs for HDI have been demonstrated to be feasible options for scenarios that require the operator and the drone to have a higher level of communication, thus expanding the range of applications to include human-drone partnership tasks.

Future work in this line of research will target implementing the proposed interaction techniques on Human-Multi-Drone Interaction. Here, users will have the ability

[6]Online: `http://www.madrimasd.org/lanochedelos\investigadores/actividad/vuelo-de-drones-y-m%C3%A1s-robots-asombrosos?lan=en`

[7]Online: `http://www.escuelaindustrialesupm.com/escuela-industriales-upm/ciencia-y-discapacidad-\erase-una-vez-un-parque`

to choose from different ways of interacting with multiple drones simultaneously. This type of interaction could be useful for applications such as: operator guidance of a fleet of autonomous air tankers used for aerial firefighting or massive search-and-rescue missions where one person or a team can interact with multiple drones at the same time to aid in aerial imaging reconnaissance.

REFERENCES

[1] V. Alvarez-Santos, R. Iglesias, X. M. Pardo, C. V. Regueiro, and A. Canedo-Rodriguez. Gesture-based interaction with voice feedback for a tour-guide robot. *J. Vis. Comun. Image Represent.*, 25(2):499–509, February 2014.

[2] ETH Zürich Flying Machine Arena. (2011, jul 2) *Controlling a Quadrotor Using Kinect* [online]. available: http://spectrum.ieee.org/automaton/robotics/robotics-software/quadrotor-interaction.

[3] Thomas Baudel and Michel Beaudouin-Lafon. Charade: Remote control of objects using free-hand gestures. *Commun. ACM*, 36(7):28–35, July 1993.

[4] Richard A. Bolt. *Put-that-There*: Voice and gesture at the graphics interface. *SIGGRAPH Comput. Graph.*, 14(3):262–270, July 1980.

[5] Tomás Mantecón del Valle, Carlos Roberto del Blanco Adán, Fernando Jaureguizar Núñez, and Narciso García Santos. New generation of human machine interfaces for controlling uav through depth based gesture recognition. In *Proceedings of SPIE Defense, Security and Sensing Conference*, volume 9084, May 2014.

[6] G. Dudek, J. Sattar, and Anqi Xu. A visual language for robot control and programming: A human-interface study. In *Robotics and Automation, 2007 IEEE International Conference on*, pages 2507–2513, April 2007.

[7] S.S. Fels and G.E. Hinton. Glove-talk: a neural network interface between a data-glove and a speech synthesizer. *Neural Networks, IEEE Transactions on*, 4(1):2–8, Jan 1993.

[8] M. Fiala. Artag, a fiducial marker system using digital techniques. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 590–596 vol. 2, June 2005.

[9] S. Garrido-Jurado, R. Muoz-Salinas, F.J. Madrid-Cuevas, and M.J. Marn-Jimnez. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*, 47(6):2280 – 2292, 2014.

[10] Eberhard Graether and Florian Mueller. Joggobot: A flying robot as jogging companion. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, pages 1063–1066, New York, NY, USA, 2012. ACM.

[11] Jihyun Han and Nicolas Gold. Lessons learned in exploring the leap motion(tm) sensor for gesture-based instrument design. In Baptiste Caramiaux, Koray Tahiroglu, Rebecca Fiebrink, and Atau Tanaka, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 371–374, London, United Kingdom, June 30 – July 03 2014. Goldsmiths, University of London.

[12] Chao Hu, M.Q. Meng, P.X. Liu, and Xiang Wang. Visual gesture recognition for human-machine interface of robot teleoperation. In *Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, volume 2, pages 1560–1565 vol.2, Oct 2003.

[13] G. Jones, N. Berthouze, R. Bielski, and S. Julier. Towards a situated, multimodal interface for multiple uav control. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 1739–1744, May 2010.

[14] Myung-Ho Ju and Hang-Bong Kang. Human robot interaction using face pose recognition. In *Consumer Electronics, 2007. ICCE 2007. Digest of Technical Papers. International Conference on*, pages 1–2, Jan 2007.

[15] Martin Kaltenbrunner and Ross Bencina. reactivision: A computer-vision framework for table-based tangible interaction. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction*, TEI '07, pages 69–74, New York, NY, USA, 2007. ACM.

[16] Manju Kaushik and Rashmi Jain. Gesture based interaction NUI: an overview. *CoRR*, abs/1404.2364, 2014.

[17] K.-F. Lee, H.-W. Hon, and R. Reddy. An overview of the sphinx speech recognition system. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 38(1):35–45, Jan 1990.

[18] S. Lupashin, A. Schollig, M. Sherback, and R. D'Andrea. A simple learning strategy for high-speed quadrocopter multi-flips. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 1642–1648, May 2010.

[19] A. Mashood, H. Noura, I. Jawhar, and N. Mohamed. A gesture based kinect for quadrotor control. In *Information and Communication Technology Research (ICTRC), 2015 International Conference on*, pages 298–301, May 2015.

[20] Gerard Medioni and Sing Bing Kang. *Emerging Topics in Computer Vision*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2004.

[21] V.M. Monajjemi, J. Wawerla, R. Vaughan, and G. Mori. Hri in the sky: Creating and commanding teams of uavs with a vision-mediated gestural interface. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 617–623, Nov 2013.

[22] Florian 'Floyd' Mueller and Matthew Muirhead. Jogging with a quadcopter. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2023–2032, New York, NY, USA, 2015. ACM.

[23] M. Muller, S. Lupashin, and R. D'Andrea. Quadrocopter ball juggling. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 5113–5120, Sept 2011.

[24] Jakob Nielsen. Noncommand user interfaces. *Commun. ACM*, 36(4):83–99, April 1993.

[25] V.I. Pavlovic, R. Sharma, and T.S. Huang. Visual interpretation of hand gestures for human-computer interaction: a review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):677–695, Jul 1997.

[26] J. Pestana, J.L. Sanchez-Lopez, S. Saripalli, and P. Campoy. Computer vision based general object following for gps-denied multirotor unmanned vehicles. In *American Control Conference (ACC), 2014*, pages 1886–1891, June 2014.

[27] P. Phamduy, M. DeBellis, and M. Porfiri. Controlling a robotic fish via a natural user interface for informal science education. *Multimedia, IEEE Transactions on*, 17(12):2328–2337, Dec 2015.

[28] David Vissiere Nicolas Petit Pierre-Jean Bristeau, Francois Callou. The navigation and control technology inside the ar.drone micro uav. In *18th IFAC World Congress*, pages 1477–1484, Milano, Italy, 2011.

[29] Michael I Posner, Mary J Nissen, and Raymond M Klein. Visual dominance: an information-processing account of its origins and significance. *Psychological review*, 83(2):157, 1976.

[30] Jenny Preece, Yvonne Rogers, and Helen Sharp. *Beyond Interaction Design: Beyond Human-Computer Interaction*. John Wiley & Sons, Inc., New York, NY, USA, 2001.

[31] M. Quigley, M.A. Goodrich, and R.W. Beard. Semi-autonomous human-uav interfaces for fixed-wing mini-uavs. In *Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 3, pages 2457–2462 vol.3, Sept 2004.

[32] Dilman Salih. *Natural User Interfaces*. Research Topics in HCI, School of Computer Science, University of Birmingham, Birmingham, 2015.

[33] J.L. Sanchez-Lopez, R. Suarez-Fernandez, H. Bavle, C. Sampedro, M. Molina, and P. Campoy. Aerostack: An architecture and open-source software framework for aerial robotics. In *Unmanned Aircraft Systems (ICUAS), 2016 International Conference on*, page 0, June 2016.

[34] Jose Luis Sanchez-Lopez, Jesús Pestana, Paloma Puente, and Pascual Campoy. A reliable open-source system architecture for the fast designing and prototyping of autonomous multi-uav systems: Simulation and experimentation. *Journal of Intelligent & Robotic Systems*, pages 1–19, 2015.

[35] Andrea Sanna, Fabrizio Lamberti, Gianluca Paravati, and Federico Manuri. A kinect-based natural interface for quadrotor control. *Entertainment Computing*, 4(3):179 – 186, 2013.

[36] Matthew Turk and George Robertson. Perceptual user interfaces (introduction). *Commun. ACM*, 43(3):32–34, March 2000.

[37] Daniel Wigdor and Dennis Wixon. *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1st edition, 2011.