

Methods for Autonomous Wristband Placement with a Search-and-Rescue Aerial Manipulator

Jesús M. Gómez-de-Gabriel, Juan M. Gandarias, Francisco J. Pérez-Maldonado
Francisco J. García-Núñez, Emilio J. Fernández-García and Alfonso J. García-Cerezo
System Engineering and Automation Department
University of Málaga
Málaga, Spain
jesus.gomez@uma.es

Abstract—A new robotic system for Search And Rescue (SAR) operations based on the automatic wristband placement on the victims’ arm, which may provide identification, beaconing and remote sensor readings for continuous health monitoring. This paper focuses on the development of the automatic target localization and the device placement using an unmanned aerial manipulator. The automatic wrist detection and localization system uses an RGB-D camera and a convolutional neural network based on the region faster method (Faster R-CNN). A lightweight parallel delta manipulator with a large workspace has been built, and a new design of a wristband in the form of a passive detachable gripper, is presented, which under contact, automatically attaches to the human, while disengages from the manipulator. A new trajectory planning method has been used to minimize the torques caused by the external forces during contact, which cause attitude perturbations. Experiments have been done to evaluate the machine learning method for detection and location, and for the assessment of the performance of the trajectory planning method. The results show how the VGG-16 neural network provides a detection accuracy of 67.99%. Moreover, simulation experiments have been done to show that the new trajectories minimize the perturbations to the aerial platform.

I. INTRODUCTION

First response teams play a crucial role when natural or man-made disasters occur. Specific preparedness are specially important to reduce the damage caused by these events. However, when disasters arise suddenly is difficult to elaborate an actuation plan for minimizing the harm [1]. Therefore, most common response to these types of incidents are usually addressed by rescuing victims and transferring them to a hospital. In these Search And Rescue (SAR) operations, a rapidly search and triage of the victims is the main goal.

Dealing with the monitoring of a large number of casualties is key to an effective response in disaster situations. First responders have to rapidly triage injured in a coordinated manner to prevent the overwhelming of the emergency field and hospital staff [2]. As the conditions of the victims change over time, remote monitoring methods are helpful. In [3] a vital signs sensor that straps to the victim’s chest has been proposed, that collect several vital signs, including respiratory activity.

With the use of robots, localization an continuous monitoring is possible by attaching sensors to the victims for identification and status assessment. In [4], a sensorized drone can reach the victim providing a quick evaluation of the

victim’s status to help rescue services. In [5] a robotic sensor has been attached to a victim that allow remote echography. Also, robots with two-way multimedia communication, for the interaction with trapped victims, have been developed [6]. However, these solutions keep the robot busy during the monitoring, so for massive disaster scenarios, a more distributed solution is needed.

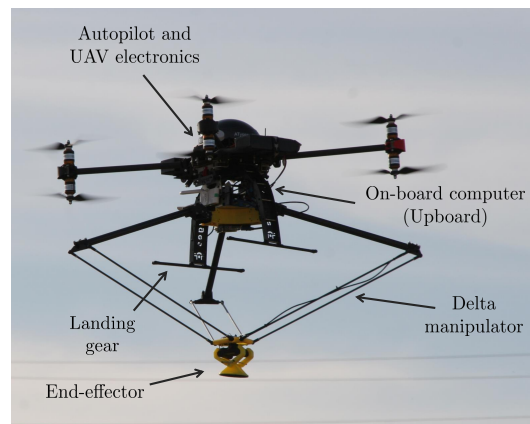


Fig. 1. Aerial manipulator based on a commercial ATyges octocopter with an extended landing gear and a delta manipulator with large workspace.

Unmanned Aerial Vehicles (UAVs) are becoming common in SAR robotic teams. They are capable of carrying out increasingly complex missions with high levels of autonomy [7]. The resurgence of artificial intelligence, specially deep learning-based techniques, is providing powerful tools for SAR operations [8]. In [9], a machine learning method is trained using positive and negative images of human bodies to locate victims from UAVs. Also, the popularity of RGB-D cameras help robots to localize and detect objects [10].

An aerial manipulator combines the advantages of UAVs with the versatility of manipulators as a single device. They range from conventional serial manipulators [11] to parallel manipulators (as in Figure 1) which take advantage of limited center of mass displacements [12] and even with the combination of multiple multi-rotors [13] for increased manipulability. The compliant interaction with the environment in aerial manipulators is a current research topic, not only with single [14] but also with dual arms [15], [16]. The design of aerial

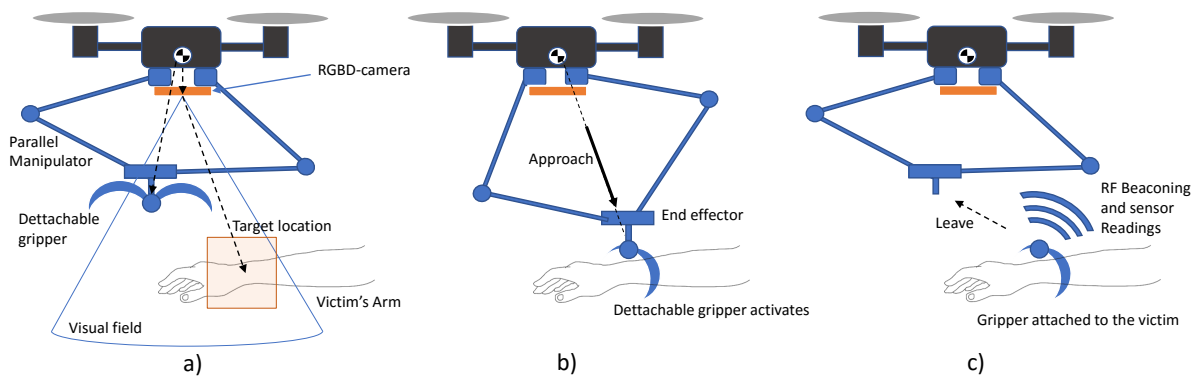


Fig. 2. Summary of the wristband placement task. a) Detection: an RGB-D camera and a Machine Learning method provides Cartesian target coordinates. b) Approach: Motion along a radial trajectory towards the target position. The gripper automatically closes around the human arm and detaches form the manipulator. c) Depart: The aerial manipulator moves away form the human.

manipulators for autonomous human interaction SAR must solve the problems of the intelligent perception and safe actuation, keeping the humans – and their limbs – away from the propellers of the aerial platform. This way, a manipulator with extended workspace is needed and the physical Human-Robot Interaction (pHRI) minimized as in [17], where a simple contact drone-wall compliance is studied.

The presented work is based on the development of an aerial manipulators for autonomous wristband placement in victims, that could send useful information for SAR operations. The system uses Machine Learning methods to detect an locate human wrists, and a parallel manipulator with an special trajectory generation method that minimizes the effects of the contact reaction forces. The pHRI is kept to a short contact operation, thanks to an automatically detachable gripper (wristband). After the contact, the aerial manipulator is dismissed, and can be used again with another victim. The main contribution of this paper is the implementation and evaluation of methods for robotic autonomous operation in SAR robots, that include: the application of the Faster R-CNN method for the visual localization of human wrists with an RGB-D camera; the design of a new device, deployable from UAVs that can carry human monitoring sensors; and a trajectory generation method for a parallel manipulator that minimizes the effects of the contact reaction forces.

This paper is structured as follows: Section II describes the proposed task and the components of the robotic system. Section III presents the method for the automatic human wrists detection and localization. In section IV, the trajectory generation method is presented. The experiments and results are included in section V, and finally, section VI offers the conclusions.

II. SYSTEM DESCRIPTION

A. Wristband Placement Task

The wristband placement task can be summarized as shown in Figure 2 as three main stages:

- 1) Detection: The human wrist is detected and located inside the visual area of the RGB-D camera, providing

the Cartesian coordinates of the target (human wrist) if found (Figure 2a).

- 2) Approach: The manipulator then moves from the resting position to the target position along a suitable trajectory, with a fast motion profile. The contact force activates a gripper detaching mechanism (Figure 2b).

- 3) Depart: The aerial manipulator is free. It moves away from the human, and a the manipulator gets back to landing position with a slow motion profile (Figure 2c).

Next, the wristband may provide different functions like beaconing, identification, and remote sensor reading for vital signs monitoring (Pulse rate, blood oxygen perfusion, temperature, accelerometer, ...) but it is not in the scope of this work.

This autonomous operation is part of a SAR mission, where the exploration and human body location tasks can be performed in teleoperated or autonomous manner by other team members.

B. The Detachable Gripper

The wristband has been designed and built in the form of a double effect detachable gripper, as can be seen in Figure 3. The gripper is passive (doesn't require power) and uses an extension spring to keep the mechanism in one of the two stable states. A prototype has been built using 3D printing and the design has been made available on-line¹.

This device acts as a carrier for the desired monitoring electronics that can be battery powered. When the wristband (gripper) has been attached, the aerial manipulator sends the GPS coordinates and the device number to the mission control system.

C. Multi-Copter

A heavy duty Drone from ATyges (Málaga, Spain) has been chosen as the aerial platform. Weights 8.4kg plus 1.2kg of batteries, and has about 30 minutes of flying time. The landing gear has been enlarged to accommodate the delta manipulator, as can be seen in Figure 1.

¹<https://www.thingiverse.com/thing:2996502>

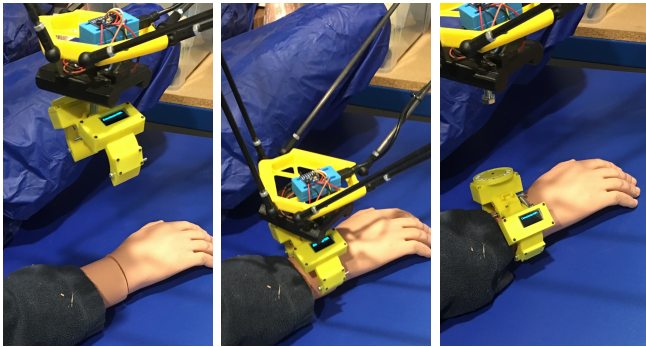


Fig. 3. The passive detachable gripper showing the approximation (left), the contact (middle) and the release (right) stages on a dummy’s wrist.

D. Parallel Manipulator

A delta-type parallel manipulator has been built. With just three degrees of freedom and parallel structure it keeps the center of mass near the base. To make it lightweight, carbon fiber tubing for the links, and XT-CF20 (Nylon with 20% of Carbon Fiber) 3D printing material has been used for the joints. Off-the-shelf plastic bearings and ABS has been used for other parts.

The manipulator weight 1.15kg (including motors), its workspace is a cone of about 1.1m high and is driven with three smart actuators Dynamixel MX-64R from Robotis, with integrated PID controller.

E. Electronics

The main on-board computing hardware in the system is an embedded computer UP-Board (Intel) with Quad Core x86, 64 bit architecture running Linux OS. A USB Wireless stick has been used for development, and a serial to USB interface for the Dynamixel servos have been added.

To power the computer and servos from the main drone batteries, two 5A DC-DC converters have been used.

F. Software

The control software has been implemented in ROS² to facilitate the integration and communication between the different modules. ROS Modules for the RGB-D camera and point cloud handling have been obtained. The Faster R-CNN has been developed in the Caffe Framework. The rest of the software has been programmed using the Python language.

III. HUMAN WRIST LOCALIZATION

A deep learning method has been implemented for the image-based wrist detection. The layout of the method is presented in Fig. 4. First of all, from a sequence of images taken by the RGB-D camera, each RGB image \mathcal{I} ($256 \times 256 \times 3$) is the input for the *Faster R-CNN*. This network gives the region coordinates in pixels with respect to the reference system of the image plane $[x_{c_0}, y_{c_0}, x_{c_1}, y_{c_1}]$ and the accuracy estimation of the object classification \mathcal{L} .

Then, the depth map \mathcal{D} taken by the RGB-D camera along \mathcal{I} is used by the *Point Cloud Library* (PCL) to generate a

Point Cloud \mathcal{P} of the scene which is used to obtain the goal position $[x, y, z]$. Finally, the goal position and the class score are published in a 4-component vector $[x, y, z, \mathcal{L}]$.

A. Faster R-CNN

Faster R-CNN [18] is one of the most important contributions to the current state-of-the-art in object detection networks. This method consists of a deep neural network for both object classification and detection from images. The primary objective of that work was to share computation with the *Fast R-CNN* object detection network [19], which was conceived from the original *R-CNN* [20]. That way, a Region Proposal Network (*RPN*) was introduced in *Faster R-CNN*. The *RPN* takes the feature map \mathcal{F} as an input and outputs a set of rectangular images proposals with different shapes and sizes, commonly called boxes or Regions Of Interests (*ROIs*), and the classification score of the object presented in each region.

In this application the *Faster R-CNN* with the *VGG-16* [21] has been implemented. The architecture of the network is presented in Fig. 4 within the discontinuous line frame. The first convolutional layers of the *VGG-16* network outputs \mathcal{F} from \mathcal{I} . Then, \mathcal{F} is used by the *RPN* to obtain the *ROIs*. Finally, after the *ROI pooling* layer two fully connected layers (*fc6* and *fc7*) output the box coordinates and the classification scores which are converted into probabilities \mathcal{L} passing through a *softmax* layer.

B. Dataset

To train and test the proposed method, a dataset formed by 5517 samples of RGB images has been used, where 5117 regions of interest around human wrists have been detected and located in each image. The dataset has been obtained from *Hand Dataset* [22]. From this database, a wrist-specific dataset has been created using the *Training Image Labeler* tool from the *Image Processing and Computer Vision* Matlab Toolbox. Fig. 5 shows four images of the custom-made dataset where the ROI around the wrist is highlighted in red. The dataset is split out in the training and test set. The training set includes 4205 samples of RGB images with human wrists, whereas the test set has 1312 samples (912 with human wrists and 400 without human wrists).

C. Training

In order to minimize the training time of the network, a transfer learning approach has been followed using a pre-trained VGG-16 CNN, taking advantage of the fact that the convolutional layers of learns to extract features from images. The remainders layers of the *Faster R-CNN* network are trained for four days following the exponential decay policy, which consists of reducing the learning rate by steps according to (1).

$$\varepsilon_n = \varepsilon_0 \cdot \gamma^{(n/\xi)} \quad (1)$$

Where ε_n is the learning rate at the iteration n , ε_0 is the initial learning rate ($\varepsilon_0 = 0.001$), γ is the reduction parameter of the learning rate ($\gamma = 0.1$) and ξ is the step

²<http://www.ros.org/>

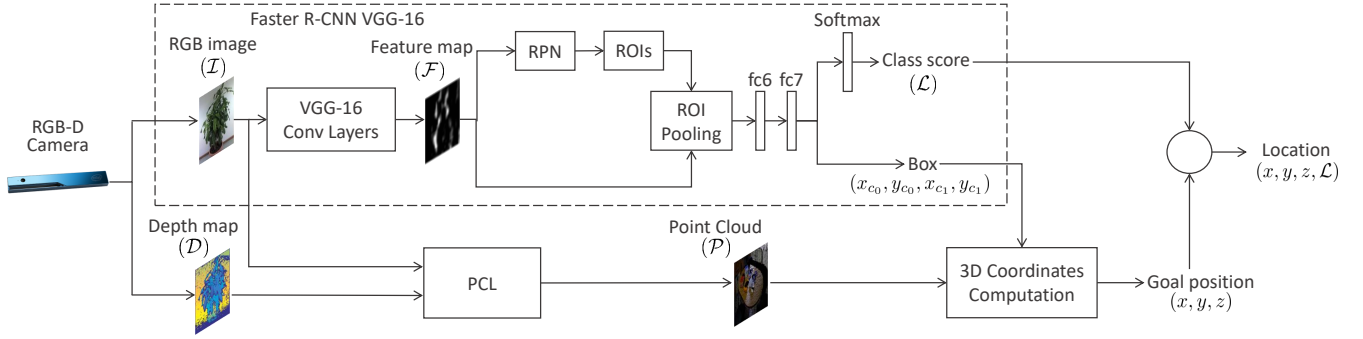


Fig. 4. Illustration of the human wrist localization algorithm. *Faster R-CNN VGG-16* outputs \mathcal{L} and $[x_{c_0}, y_{c_0}, x_{c_1}, y_{c_1}]$ form \mathcal{I} , while \mathcal{P} , generated with the *PCL*, and the box coordinates are used to compute the 3D coordinates of the goal position.



Fig. 5. Four images from the custom-made dataset used both for training and test the method.

size ($\xi = 5 \cdot 10^4$). On the other hand, the momentum (μ) has been set to 0.9. This parameters are normally chosen with similar values in other research studies [23], [24].

D. Cartesian 3D coordinates computation

After detecting a wrist in the scene, it is necessary to compute the 3D coordinates of the goal position. Taking advantage of the depth information provided by the RGB-D camera, this task can be carried out easily.

The *PCL* is used to generate \mathcal{P} from the RGB-D data. \mathcal{P} contains the information of the 3D cartesian coordinates in meters with respect to the reference system of the camera and RGB values for each pixel of the 2D \mathcal{I} . Thus, giving a target pixel (\mathcal{T}_c) from \mathcal{I} , the 3D coordinates are computed using \mathcal{P} . \mathcal{T}_c is calculated as the middle pixel of the ROI with (2).

$$\mathcal{T}_c = \left[\frac{x_{c_1} - x_{c_0}}{2}, \frac{y_{c_1} - y_{c_0}}{2} \right] \quad (2)$$

To determine the target position with respect to the reference system of the aerial manipulator, an homogeneous transformation that relates the reference system of the camera and the reference system of the manipulator is applied.

IV. TRAJECTORY PLANING

A new method has been designed to minimize the effects of the interaction forces on the aerial manipulator at the second stage of the wristband placement task (See Figure 2b). Reaction forces at the contact point are converted into forces and torques around the center of mass of the aerial manipulator. Force produce displacement disturbances to the flight control, in opposite direction to the human. In the other hand, torque produce attitude disturbances, which poses a stability problem in the proximity of a human. If the reaction force at the contact point is aligned with the center of mass, the torques are zero. Considering that the direction of the reaction force depends on the direction of the contact motion, the use of approach trajectories radial with respect to the instantaneous center of mass produces a safer and controllable disturbance.

A. Center of Mass Computation

To compute the center of mass of the entire system, the center of mass of the parallel manipulator and the UAV are computed separately. Equation (3) shows the expression for the center of mass of the parallel manipulator CM_m , which is needed to get the center of mass of the whole system CM as in (4).

$$CM_m = \sum_{i=1}^3 (\vec{r}_{a_i} \cdot m_a + \vec{r}_{e_i} \cdot m_e + \vec{r}_{f_i} \cdot m_f) \quad (3)$$

$$CM = \frac{CM_m + \vec{r}_h \cdot m_h + \vec{r}_b \cdot m_b + \vec{r}_d \cdot m_d}{3(m_a + m_e + m_f) + m_e + m_b + m_d} \quad (4)$$

Where \vec{r}_{a_i} , \vec{r}_{e_i} , \vec{r}_{f_i} , \vec{r}_e , \vec{r}_b and \vec{r}_d are the cartesian positions of the center of mass of each arm, elbow, forearm, end effector, base and UAV respectively. And m_a , m_e , m_f , m_h , m_b and m_d are the masses of the arm, elbow, forearm, end effector, the base and the UAV respectively.

B. Trajectory Generation

Two paths are proposed for every displacement to the target point. The manipulator initial position P_{Inter} , is assumed in an horizontal reference plane near the center of mass. Then, a second point is calculated in the reference plane that aligns the center of mass with the goal position and a path is generated that adapts the end effector to the next path from that point. Next, an extension path of the manipulator is generated from the previous point to the goal position radially to the instantaneous center of mass. Both paths are defined by a linear interpolation of N points, for which the center of mass is evaluated, to ensure that the path is radial at the moment of the contact, as follows:

Algorithm 1 Calculation of the center of mass

```

for i= 1 to N do
  Q[i]= InverseKinematics(X[i])
  CM= CenterofMass(Q(i))
  P_Inter= Intersection(CM, Ref_Plane, X[i])
  X[i+1]= NextPoint(P_Inter, X[i], N-1)
end for

```

To find the initial point of the contact path, a method of successive approximations is used, foreseeing possible displacements of the center of mass of the aerial manipulator:

Algorithm 2 Trajectory generation

```

P_Inter=Intersection(CM, Ref_Plane, Goal_X)
while P_Inter != X do
  X= P_Inter
  Q= InverseKinematics(X)
  CM= CenterofMass(Q[i])
  P_Inter= Intersection(CM, Ref_Plane, Goal_X)
end while

```

Finally, the Cartesian trajectory is sent to the real-time motion controller that sends joint references to the servos in sequence.

V. EXPERIMENTS AND RESULTS

A. Wrist Localization

To measure the human wrists detection accuracy of the network, an experiment with 1312 samples of RGB images, including 912 with wrists and 400 without wrists, has been carried out. This experiment shows the performance of the network detecting the presence of a wrist when it is present in the scene and detecting the lack of wrists present in a scene where no wrist is present. Cases in which a wrist is present in an image and the network detects it, or if there is no wrist in the image and the network does not detect it, are considered successful.

The experiment results are shown in Table I. The results exhibit an accuracy of 57.13% detecting wrists when they are in the images, and an accuracy of 92.75% when there are no wrists in the images with different backgrounds and lighting conditions. The Average Precision (AP) measures the

TABLE I
WRIST DETECTION EXPERIMENTS

	Wrist detected	No wrist detected
Wrist presented	521(57.13%)	391(42.87%)
No wrist presented	29(7.25%)	371(92.75%)

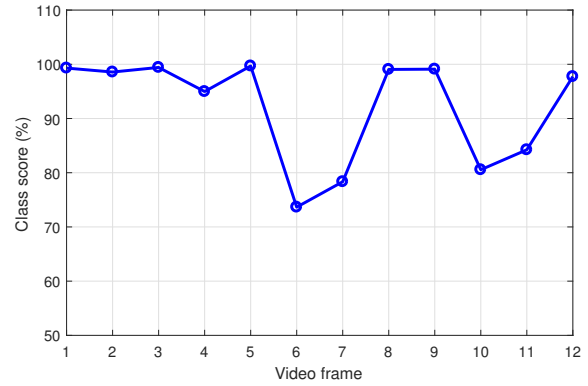


Fig. 6. Class score achieved in each frame of the sequence of images after passing through the *Faster R-CNN*.

average of the positives cases. That way, an AP of 67.99% has been achieved by the detection method.

Moreover, an experiment to evaluate the performance of the localization method has been carried out. A sequence of images with 12 frames has been considered. A human wrist is present in every frame. Fig. 6 shows the class score achieved in each frame after passing through the *Faster R-CNN VGG-16*. The first 8 images of the sequence is shown in Fig. 7, where the ROI and the achieved class score are also highlighted. On the other hand, Fig. 8 presents two images where a human wrist is present in different positions. The 4-component vector on the images corresponds the output given by the localization method that contains the position and the class score.

B. Detachable Gripper Experiments

A prototype of the detachable gripper has been built and a video of the wristband placement experiments has been made available on-line³ to show the operation of this new device. The device worked as expected.

C. Trajectory Planning

A dynamic simulation of the aerial manipulator has been implemented to evaluate the effects of the load and external reaction forces over the center of the aerial platform, using the new radial trajectory generation method, and non-radial trajectories.

The experiment in Figure 9, shows that the use of an arbitrary approaching trajectory, creates disturbances in the form of forces and torques around the center of mass. On the other hand, The experiment in Figure 10, shows that the use of a trajectory radial to the center of mass, as expected, minimizes the torques at the platform.

³<https://youtu.be/dMuaDy22YuU>

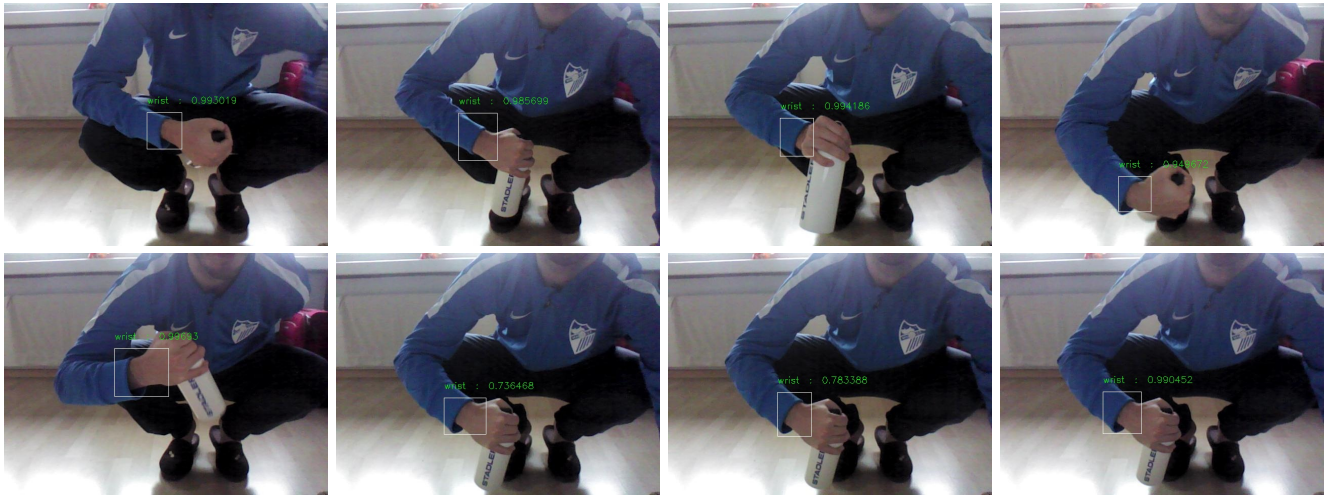


Fig. 7. Frames 1 to 8 (from left to right and top to bottom) of the sequence of images used for the wrist detection experiment. In each frame, a class score $\mathcal{L} > 90\%$ is achieved.

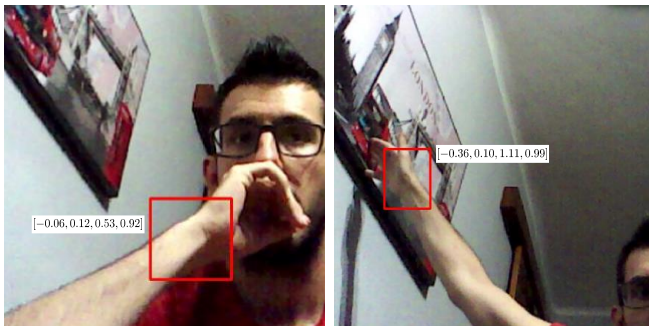


Fig. 8. Localization of the wrist in two different positions, showing the pose obtained with respect to the reference system of the camera and the class score in the detection vector $[x, y, z, \mathcal{L}]$. Note that in both pictures $\mathcal{L} > 90\%$.

VI. CONCLUSIONS

In this paper, a new system for the autonomous wrist placement using an aerial manipulator, with applications to SAR operations has been presented. A new task has been defined in which the UAV autonomously places a wristband on the victims, with a new design that automatically transfers between the manipulator and the human arm. This wristband can be used as a carrier for other equipments, opening the door to many new remote monitoring applications. A machine learning method has been successfully applied, and evaluated in the autonomous detection and localization of human wrists, based on the Faster R-CNN. Experiments to test the detection accuracy of the network has been carried out, showing an AP of 67.99%. A new trajectory planning method for aerial manipulators that minimizes the effects produced by the contact reaction forces on the flight control has been presented. The results of the experiments have shown that with the new method, the dynamic effects of the environment compliance are minimized providing a safer and stable control of the UAV.

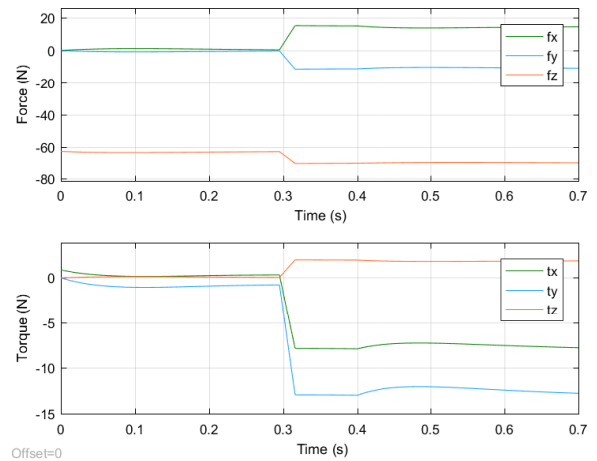


Fig. 9. Dynamic simulation of the effects of external reaction forces measured at the center of the aerial platform, with a no radial trajectory. External forces are applied for 0.1s in the opposite direction to the end effector speed, showing high-torque perturbations.

ACKNOWLEDGMENT

This work was funded by the Spanish project DPI2015-65186-R , the European Commission under grant agreement BES-2016-078237 and the University of Malaga.

REFERENCES

- [1] Ciro Ugarte, Jacobo A. Tieffenberg, Ribka Amsalu, Lou E. Romig, and Tien T. Vu. Planning and triage in the disaster scenario. *American Academy of Pediatrics, Pediatric Education in Disasters Manual, Module 3*.
- [2] Eric Frykberg and Joseph Tepas. Terrorist bombings. lessons learned from belfast to beirut. *Annals of surgery*, 208(5):569, 1988.
- [3] G. C. Pallis, N. Ferreira, L. Hildebrand, and G. Seynaeve. Wireless transmission of vital signs of entrapped victims during search and rescue operations in collapsed buildings. In *4th International Conference on Wireless Mobile Communication and Healthcare(MOBIHEALTH)*, pages 254–257, Nov 2014.
- [4] A. J. A. Dhivya and J. Premkumar. Quadcopter based technology for an emergency healthcare. In *Third International Conference on Biosignals, Images and Instrumentation (ICBSII)*, pages 1–3, 2017.

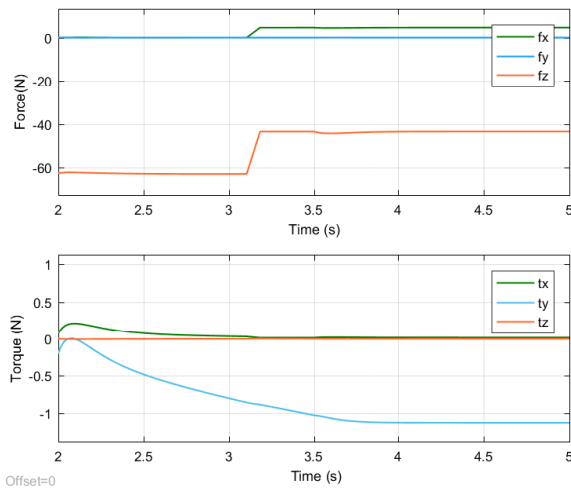


Fig. 10. Dynamic simulation of the effects of a load and external reaction forces measured at the center of the aerial platform, using the new radial trajectory method. External forces are applied for 0.1s in the opposite direction to the end effector speed and the showing negligible torque perturbations due to the external reaction force.

[5] K. Ito, S. Sugano, and H. Iwata. Development of attachable tele-echography robot by a bystander at injury scene. In *IEEE International Conference on Mechatronics and Automation*, pages 1270–1275, 2010.

[6] R. R. Murphy, V. Srinivasan, Z. Henkel, J. Suarez, M. Minson, J. C. Straus, S. Hempstead, T. Valdez, and S. Egawa. Interacting with trapped victims using robots. In *IEEE International Conference on Technologies for Homeland Security (HST)*, pages 32–37, 2013.

[7] Iván Maza, Fernando Caballero, Jesús Capitán, José Ramiro Martínez-de Dios, and Aníbal Ollero. Experimental results in multi-uav coordination for disaster management and civil security applications. *Journal of intelligent & robotic systems*, 61(1-4):563–585, 2011.

[8] Mesay Bejiga, Abdallah Zeggada, Abdelhamid Nouffidj, and Farid Melgani. A convolutional neural network approach for assisting avalanche search and rescue operations with uav imagery. *Remote Sensing*, 9(2):100, 2017.

[9] M. T. Rashid and S. A. Mahmud. Intelligent seeker robot to locate dead bodies and survivors under debris. In *IEEE Global Humanitarian Technology Conference (GHTC)*, pages 1–6, 2017.

[10] Ian Lenz, Honglak Lee, and Ashutosh Saxena. Deep learning for

detecting robotic grasps. *The International Journal of Robotics Research*, 34(4-5):705–724, 2015.

[11] X. Meng, Y. He, F. Gu, L. Yang, B. Dai, Z. Liu, and J. Han. Design and implementation of rotor aerial manipulator system. In *IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 673–678, 2016.

[12] T. W. Danko, K. P. Chaney, and P. Y. Oh. A parallel manipulator for mobile manipulating uavs. In *IEEE International Conference on Technologies for Practical Robot Applications (TePRA)*, pages 1–6, 2015.

[13] J. Mendoza-Mendoza, G. Sepulveda-Cervantes, C. Aguilar-Ibanez, M. Mendez, M. Reyes-Larios, P. Matabuena, and J. Gonzalez-Avila. Air-arm: A new kind of flying manipulator. In *Workshop on Research, Education and Development of Unmanned Aerial Systems (RED-UAS)*, pages 278–287, 2015.

[14] A. E. Jimenez-Cano, G. Heredia, and A. Ollero. Aerial manipulator with a compliant arm for bridge inspection. In *International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 1217–1222, 2017.

[15] A. Suarez, G. Heredia, and A. Ollero. Physical-virtual impedance control in ultra-lightweight and compliant dual arm aerial manipulators. *IEEE Robotics and Automation Letters*, PP(99), 2018.

[16] V. Lippiello, G. A. Fontanelli, and F. Ruggiero. Image-based visual-impedance control of a dual-arm aerial manipulator. *IEEE Robotics and Automation Letters*, PP(99), 2018.

[17] T. Bartelds, A. Capra, S. Hamaza, S. Stramigioli, and M. Fumagalli. Compliant aerial manipulators: Toward a new generation of aerial robotic workers. *IEEE Robotics and Automation Letters*, 1(1):477–483, 2016.

[18] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.

[19] Ross Girshick. Fast r-cnn. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1440–1448. IEEE, 2015.

[20] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 580–587. IEEE, 6 2014.

[21] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv*, 2014.

[22] A. Mittal, A. Zisserman, and P. H. S. Torr. Hand detection using multiple proposals. In *British Machine Vision Conference*, 2011.

[23] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015.

[24] Samuel L. Smith, Pieter-Jan Kindermans, and Quoc V. Le. Don't decay the learning rate, increase the batch size. *arXiv preprint arXiv:1711.00489*, 2017.