

Saliency-Based Cooperative Landing of a Multirotor Aerial Vehicle on an Autonomous Surface Vehicle

João Silva, Ricardo Mendonça, Francisco Marques, Paulo Rodrigues, Pedro Santana and José Barata

Abstract—This paper presents a method for vision-based landing of a multirotor unmanned aerial vehicle (UAV) on an autonomous surface vehicle (ASV) equipped with a helipad. The method includes a mechanism for helipad behavioural search when outside the UAV's field of view, a learning saliency-based mechanism for visual tracking the helipad, and a cooperative strategy for the final vision-based landing phase. Learning how to track the helipad from above occurs during takeoff and cooperation results from having the ASV tracking the UAV for assisting its landing. A set of experimental results with both simulated and physical robots show the feasibility of the presented method.

I. INTRODUCTION

Due to their wide application range, unmanned aerial vehicles (UAV) are attracting considerable attention of the robotics community. This is particularly striking for multirotor vehicles, which are capable of vertical takeoff and landing. Despite all the developments on this domain, landing these vehicles outdoors on a dynamic landing platform, helipad hereafter, is still an open problem.

This paper addresses the particular problem of devising a end-to-end search and landing strategy for a UAV that needs to return to its helipad (see Fig. 1), even when the helipad is not covered by the UAV's downwards looking camera's field of view and its position is highly uncertain to the UAV. To solve this hard problem, the UAV needs to infer the vantage point that most probably will bring the helipad into its field of view and then to visually detect and track the helipad while descending.

Detecting and tracking a helipad from high altitudes is rather difficult given its small size. To diminish this problem, the UAV learns the altitude-dependent appearance of the helipad during takeoff. To also consider how the helipad differs from the background, the learning process occurs over a computational model of visual saliency.

The helipad is assumed to be mounted on a mobile robot, concretely, an autonomous surface vehicle (ASV) running an environmental monitoring task. This means that the helipad can be seen as a mechatronic device capable of perception, reasoning, and communicating with other devices, such as the UAV. These assets are herein exploited in order to allow the helipad to assist the UAV in the search and land task.

The helipad assists the search by providing the UAV with estimates of their relative pose. This information is as

P. Santana is with ISCTE - Instituto Universitário de Lisboa (ISCTE-IUL), Portugal and Instituto de Telecomunicações (IT), Portugal. E-mail: pedro.santana@iscte.pt

J. Silva, R. Mendonça, F. Marques, P. Rodrigues, and J. Barata are with CTS-UNINOVA, Universidade Nova de Lisboa (UNL), Portugal.



Fig. 1. The hexacopter used in the experiments. This UAV is equipped with a marker to be tracked from the helipad's upwards looking camera and a downwards looking camera to track the helipad. Both marker and UAV camera are mounted on a mechanical gimbal.

inaccurate and seldom as higher is the distance between the UAV and the helipad. As a result, the UAV needs to exploit the pose information as a cue in a context-sensitive fusion process. When the distance is short, as it is a few moments before landing, the communications become reliable and the assistance provided by the helipad can be used within the UAV's control loop. This is exploited by letting the helipad to visually track the UAV with an upwards looking camera and report their relative accurate pose.

This paper is organised as follows. Section II discusses related work. Then, an overview of the system is given in Section III. The global search process that allows the UAV to search the helipad, when it is not visible, is presented in Section IV. Subsequently, the local search phase, i.e., when the helipad is visible to the UAV, is presented in Section V. The system description is finalised with the landing process in Section VI. A set of experimental results in simulation and with physical robots is provided in Section VII. Finally, conclusions and future work avenues are discussed in Section VIII.

II. RELATED WORK

The presence of a human operator can facilitate the search and landing problem. For instance, the landing target can be defined by a GPS position that is to be tracked by the UAV [1]. Rather than a GPS position, the human operator may define the landing target on an image, which is then used as a template for visual search of the landing site. This has been done using for instance infrared imagery [2].

Not relying on a human operator is pivotal when communications are unavailable. In this case, the UAV may need to engage on an emergency safe landing, which means

picking the portion of the terrain that poses the least risk to the landing process, given the tridimensional environment information gathered online [3]. The recent developments in robust tridimensional reconstruction from aerial platforms (e.g., [4]) will be key to push safe landing in natural environments forward and to enable human-directed landing on known a priori structures in the environment.

Landing on a pre-specified docking station, such as a helipad, is key to allow UAVs to work together with other robots. See for instance the case of having UAVs and ASVs cooperating for the accomplishment of environmental monitoring tasks [5]. Moreover, the use of a helipad reduces considerably the computational effort, as compared to solutions that demand tridimensional reconstruction from vision sensors.

Helipads may exhibit different appearances, each providing different accuracy-robustness-speed trade-offs. The presence of a texture in the landing site has been shown to be sufficient to guide the landing process via optic flow estimation [6]. However, the presence of a well defined helipad reduces computation and fosters accuracy.

Most often helipads exhibit black/white visual patterns that can be easily detected from aerial images, such as chessboards [7], white on black squares [8], concentric rings [9], and H-based patterns [10]. In addition to be able to detect the helipad, it is also important to be able to track it, which is commonly done with Kalman filtering [11]. Markers as those found in augmented reality have been used as well for controlling a UAV in a multi-camera setup [12].

A limitation of these techniques alone is that they rely on visual acuity for detailed marker detection, thus limiting their applicability for higher altitudes. With a saliency-based approach, our system overcomes this limitation by bringing to the system the information related to how the helipad's appearance relates to the background's at several scales. This means learning the relationship between the appearance of the helipad and supporting mobile robot and between this assemble and the background environment. This regional descriptor is robust when the details are lost due to altitude, image deformation, and noise.

III. SYSTEM OVERVIEW

A. The Robotic System

The robotic system is composed of an hexa-rotor UAV with vertical takeoff and landing capabilities and a helipad mounted on a mobile robot. Both robots are provided with a GPS and an IMU for pose estimation in outdoor environments. Both robots are able to reliably communicate with each other via a wireless bidirectional communication channel only in the short range, i.e., below 100m. At greater distances both robots will eventually lose communications. The helipad has an upwards looking camera attached to its centre. This camera will allow the helipad to assist the UAV in the final landing phase. The UAV is equipped with a downwards looking camera whose optical axis is aligned with the gravity vector. This alignment is physically ensured by having the camera mounted on a gimbal. The helipad

is assumed to be mounted on a mobile robot moving on a planar surface always perpendicular to the gravity vector. As a result, the upwards looking camera's optical axis is parallel to the gravity vector.

B. Helipad Search

To allow the UAV to search for the helipad beyond its field of view, a global search is used (Section IV). The global search is the task of determining what is the next vantage point in the environment that provides higher chances for the UAV to detect the helipad. This is done by means of fusing several heuristically defined criteria.

For each vantage point selected by the global search, the UAV moves towards it and a local search is engaged therein (Section V). The local search aims at checking, in a prioritised way, if any of the objects present in the UAV's field of view at the vantage point picked by the global search is the helipad. If so, a landing behaviour is triggered (Section VI).

The local search starts by ranking all objects present in the UAV's field of view according to their likelihood of being the helipad (Section V-B). This likelihood is based on saliency-based knowledge about the relative appearance of the helipad, underlying mobile robot, and local environment, which was learned during the preceding takeoff (Section V-A). Then, each of the ranked objects of interest are scheduled for a sequential close-in inspection (Section V-C). This inspection means centring the UAV on the object of interest and lowering it down to an altitude at which a definitive decision about the presence of the helipad can be confidently made. If the helipad is found, then the landing behaviour is engaged. If the local search fails to detect the helipad in any of the inspected objects, the global search is reiterated and the whole subsequent processes repeated accordingly.

IV. GLOBAL SEARCH

The global search is assumed to occur at a fixed high altitude, a_h . The latitude and longitude of each putative vantage point are obtained using a multi-criteria optimisation process. For this purpose, a geo-referenced $n \times m$ grid is assumed to represent the robot's useful workspace, and whose origin is the helipad's position at takeoff time. Let $G_k(i, j, t) \in [0, 1]$, with $1 \leq i \leq n, 1 \leq j \leq m$, return the utility, according to a given criterion $k \in K$, of considering the grid cell (i, j) as the next vantage point from which the helipad will be locally searched. Grid cells are assumed to be squared and with an area that ensures full coverage of the ground imaged by the UAV's downwards-looking camera at the fixed search height, given the camera's intrinsics obtained from calibration.

To take into account all criteria in K , these are fused onto a global utility map:

$$\Psi(i, j, t) = \frac{1}{|K|} \sum_{k \in K} \omega_k G_k(i, j, t), \quad 1 \leq i \leq n, 1 \leq j \leq m, \quad (1)$$

where t refers to the current time step, ω_k is an empirically defined weight for criterion k , such that $\sum_{k \in K} \omega_k = 1$.

To determine the globally best vantage point taking into account what is known in the current moment t , the cell in $\Psi(\cdot, \cdot, t)$ with highest value, (\hat{i}, \hat{j}, t) , is found:

$$(\hat{i}, \hat{j}, t) = \arg \max_{(i, j)} \{\Psi(i, j, t)\}, \quad 1 \leq i \leq n, 1 \leq j \leq m. \quad (2)$$

The cartesian coordinates of the obtained grid cell (\hat{i}, \hat{j}) , assumed to be coherent with the UTM coordinate system, are converted to latitude and longitude coordinates before being sent to the UAV's control system as next waypoint to be tracked.

Four criteria are used to determine the best next vantage point, $K = \{v, p, l, a\}$. Criterion v states that all cells close to the current UAV's position should attract maximum attention:

$$G_v(i, j, t) = \exp \left(- \frac{\|(i, j) - \mathbf{p}(t)\|^2}{2\sigma_v^2} \right), \quad (3)$$

where $\mathbf{p}(t)$ represents the current position of the UAV in the grid coordinate system. This criterion aims at ensuring that the next visited waypoint is nearby the last (current) visited one. This is important as there is a high cost associated to moving the UAV from a position to another.

If known, the current position of the helipad should drive strongly the UAV to the corresponding grid cell. Even if the current position is unknown, previous known positions should attract the UAV, less strongly though. Awareness of the helipad is feasible in scenarios in which the helipad is mounted on a mobile robot with a communication channel to the UAV available. The attractiveness by the grid cells in which the helipad was present is governed by the following differential equation approximated by the Euler method:

$$\tau \cdot \dot{G}_p(i, j) = -G_p(i, j) + \kappa \cdot \exp \left(- \frac{\|(i, j) - \mathbf{d}\|^2}{2\sigma_v^2} \right), \quad (4)$$

where the initial value of $G_p(i, j)$ is 0, and $\kappa = 1$ in the instant the UAV receives positive information regarding the helipad's presence in a given cell, \mathbf{d} , and $\kappa = 0$ in all other moments. Basically, this formulation raises the utility of cell (i, j) whenever the position of the helipad is known to be the one of the cell and lowers the utility as time goes by, implementing a sort of forgetting mechanism.

To take the UAV to avoid previously visited positions, without hampering it completely, the following differential equation is used at each time step with the Euler method:

$$\tau \cdot \dot{G}_l(i, j) = (1 - G_l(i, j)) - \phi_l(i, j), \quad (5)$$

where the initial value of $G_l(i, j)$ is 1, $\phi_l(i, j) = 0$ in the instant the UAV reaches cell (i, j) and $\phi_l(i, j) = 1$ in all other moments. Without the forgetting factor, the UAV would not be able to find a dynamic helipad that has moved to a cell already visited by the UAV.

To unblock putative deadlock situations caused by cyclic updates or local minima resulting from the fusion of the previous criteria, criterion a values all positions randomly:

$$G_a(i, j, t) = rnd, \quad (6)$$

where rnd represents a number between 0 and 1 sampled from an uniform distribution for every tuple (i, j, t) .

These criteria induce the UAV to perform a spiral-like search behaviour around positions that are known to have been visited by the helipad. The stochastic criterion induces some variability to this global behaviour, which is key to cope with potentially dynamic behaviour of the helipad, sensor noise, and faulty actuation.

V. LOCAL SEARCH

A. Learning Helipad Appearance

The fact that the UAV takes off from the helipad can be exploited to learn its appearance and how it differs from the background's appearance. That is, the system should learn which are the visual features that turn the helipad both recognisable and salient with respect to the background. This section describes the learning process used to capture such knowledge.

Visual saliency information is commonly obtained by aggregating several maps of bottom-up contrast visual features (e.g., dark region on a bright background). The classical bio-inspired model proposed by Itti et al. [13] is herein considered as backbone for saliency computation. First, one dyadic Gaussian pyramid with eight levels is computed from the intensity channel. Two additional pyramids, also with eight levels, are computed to account for the Red-Green and Blue-Yellow double-opponency colour feature sub-channels. Each level corresponds to a given scale. Various scales are then used to create a set of on-off and off-on centre-surround maps per pyramid. These have higher intensity on those pixels whose corresponding feature differs the most from its surroundings. All centre-surround maps of a given kind, i.e., on-off or off-on, built from the intensity pyramid are resized to a common size, independently scaled in magnitude with the method recently proposed in [14], and finally averaged together to produce an aggregate intensity centre-surround feature map. Then, both on-off and off-on aggregate intensity centre-surround feature maps are scaled and averaged together to produce an intensity bottom-up conspicuity map. The same process applies to create Red-Green and Blue-Yellow conspicuity maps, which are then averaged together to produce a single colour bottom-up conspicuity map.

The bottom-up conspicuity maps need to be somehow blended in order to produce a final saliency map. Without a priori knowledge of the object being sought, this can be done at the cost of reporting several false positives induced by distractors distributed in the environment. To reduce this ambiguity, the relative importance of each aggregate centre-surround feature map contrast visual feature to the saliency map, which is computed as an weighted average of the bottom-up conspicuity maps, can be made a function of

learned knowledge about the object being sought [15], [16]. This knowledge is defined as the set of weights required for the weighted average of the conspicuity maps. This approach allows the system to exploit the visual features that turn the helipad simultaneously salient with respect to the background and recognisable.

As previously stated, the learning process occurs throughout the takeoff process. At each time step t , the helipad is tracked (see below) and its bounding box, $\mathbf{b}(t)$, used as a mask to determine which weights distribution generate a final saliency map that best correlates with the current position of the helipad in the visual field. This instantaneous weights vector, $\mathbf{w}(t)$, feeds a temporal smoothing filter to cope with noisy data (refer to [15], [16] for further details on the process that estimates $\mathbf{w}(t)$ from a learning mask):

$$\mathbf{w}^*(t) = (1 - \beta)\mathbf{w}^*(t-1) + \beta\mathbf{w}(t), \quad (7)$$

where β is the learning rate. As the perceived appearance of the helipad is likely to change as a function of the UAV's current altitude, caused for instance by specular reflections, the filtered weights vector needs to be used as a function of altitude. This way, the most appropriate weight vector, given the current altitude, can be obtained when searching for the helipad. For this purpose, at each time step, the tuple $(\hat{\mathbf{w}}(t), a(t), b(t))$ is appended in a set that is empty at the time step of takeoff onset, t_o :

$$W(t) = W(t-1) \cup \{(\mathbf{w}^*(t), \mathbf{b}(t), a(t))\}, \quad W(t_o) = \emptyset. \quad (8)$$

To track the helipad during the learning process, the helipad detector used for close-in inspection is used while in low altitude (see Section V-C). As for high altitudes the method used for the close-in inspection is no longer reliable (see Section V-C), a camshift tracker [17] is initialised on the last observed helipad's bounding box. Then, the tracker is fed with the saliency map after being modulated by $\mathbf{w}^*(t)$. As a result, the tracker will follow the saliency peak corresponding to the helipad's position, ensuring that learning can proceed for higher altitudes. As saliency is being modulated by the weights that are being learnt, the tracker remains fixed to the helipad as the UAV climbs.

B. Finding Objects of Interest

Let us assume that the learning process ends at time step $t_e > t_o$. Then, at each time step of subsequent local searches, $t > t_e$, $W(t_e)$ is used to recall the filtered weight vector, $\hat{\mathbf{w}}^*(t)$, and helipad bounding box, $\hat{\mathbf{b}}(t)$ for the current altitude, $a(t)$. This is done by simply returning the tuple whose altitude element best matches the current UAV's altitude. The recalled weight vector is then used to modulate the saliency computation so as to raise the chances that the helipad and its supporting structure is the most salient in the field of view.

At the local search onset, the saliency map, $S(t) \in [0, 1]$, is thresholded and segmented into a set of objects, $O(t)$, using a connected components operation. Some of the segmented objects are more likely to represent the helipad than others,

thus, they have different values for the search task. This value of an object is defined as its ability to be simultaneously salient and with a bounding box similar to the expected one:

$$v(o) = \alpha_s \cdot \bar{s}(t) + (1 - \alpha_s) \cdot \Phi(\hat{\mathbf{b}}(t), \mathbf{b}(o)), \quad (9)$$

where α_s is an empirically defined scalar, $\bar{s}(t) \in [0, 1]$ is the average saliency of the saliency map $S(t)$ and $\mathbf{b}(o)$ is the bounding box of object o , and

$$\Phi(\mathbf{a}, \mathbf{b}) = \frac{\min(\mathbf{a}_w, \mathbf{b}_w)}{\max(\mathbf{a}_w, \mathbf{b}_w)} \frac{\min(\mathbf{a}_h, \mathbf{b}_h)}{\max(\mathbf{a}_h, \mathbf{b}_h)} \quad (10)$$

returns the similarity between two bounding boxes, where \mathbf{a}_w and \mathbf{a}_h are the width and height of bounding box \mathbf{a} , respectively. The objects are ranked according to their value to the search task, resulting in the ordered set

$$R(t) = \{o_1, \dots, o_{|O(t)|}\}, \quad (11)$$

with $o_j \in O(t)$, $v(o_j) \geq v(o_{j+1})$, $1 \leq j \leq |O(t)|$. These objects are scheduled according to their order in $R(t)$ for subsequent close-in inspection.

C. Close-In Inspection

During a close-in inspection, two descending phases are considered. The first is the one taking the robot from the highest altitude, a_h to an intermediate altitude a_m . The second phase is the one that takes the UAV from the intermediate altitude $a_m < a_h$ to the altitude at which the UAV must decide whether it will proceed for landing or not, $a_l < a_m$. The decision altitude, a_l , can be formally determined as $f \cdot (w + e)/h$, where f is the camera's focal length (obtained from calibration), w is the width of the mobile robot transporting the helipad, e is the expected relative pose estimate error, h is the input image's height in pixels.

During the first phase, the UAV centres on the object picked from the saliency map for close-in inspection and then tracks it while progressively descending. The tracking starts by initialising a camshift tracker around the bounding box of the object under analysis. Then, the tracker is fed with the saliency map computed with the weight vector appropriate for the altitude at each time step, i.e., by consulting $W(t_e)$ given current altitude $a(t)$.

During the second phase of the descent, the low UAV's altitude results in the helipad filling a large portion of the robot's visual field. This renders visual saliency operators of little use in this case and, thus, a geometric description of the helipad is used for its detection and tracking. The helipad is assumed to be a black surface with a white H inscribed and surrounded by a white circle. Inspired by [18], a reference model, \mathbf{h}_r , is stored offline as a seven-dimensional vector containing the target's geometric invariant moments (Hu, 1962). The geometric invariant moments are fast to compute object descriptors known to be invariant to changes in scale, rotation, and translation.

Online, the system starts by searching all circles present in the image using an Hough-based approach. The goal is to find the circle circumscribing the H in the helipad. To

avoid false positives, the search is limited to circles of radius covered by the interval $[\gamma - \gamma\delta, \gamma + \gamma\delta]$, where $\gamma \approx f \cdot r/a(t)$, $a(t)$ is the UAV's current altitude, f is the camera's focal length, r is the circumscribing circle's expected radius, and δ is an empirically defined scalar. The set of circles respecting this constraint are stored in a set $C(t)$.

The geometrical moments of the input image constrained by the bounding box of each circle $c \in C(t)$, \mathbf{h}_c , are compared to the reference model in order to get a dissimilarity measure:

$$d_c = \sqrt{\sum_{i=1}^7 (\mathbf{h}_r^i - \mathbf{h}_c^i)^2}, \quad (12)$$

where \mathbf{h}^i is the i -th element of the model vector \mathbf{h} .

The circle c enclosing the appearance most similar to the expected one is

$$q(t) = \arg \min_{c \in C(t)} (d_c : d_c < \xi), \quad (13)$$

where ξ is an empirically defined similarity threshold to ensure that a circle is accepted only if it is at least somewhat similar to the model. When accepted, the dissimilarity of the best circle $q(t)$, $d_q(t)$, is appended to a set that is empty at the time step of descent onset, $t_d > t_e$:

$$D(t) = D(t-1) \cup \{d_q(t)\}, \quad D(t_d) = \emptyset. \quad (14)$$

When reached the altitude a_l , the close-in inspection behaviour must terminate with a decision about whether the helipad is actually present in the UAV's visual field. If so, then the landing process may proceed, otherwise the higher level local search behaviour must pick a new location for close-in inspection. The decision is based on the average similarity between observations and the amount of successful matchings:

$$(1 - \alpha)\bar{d}(t) + \alpha|D(t)|/(t - t_d) > 0.5, \quad (15)$$

where α is an empirically defined scalar that weighs the importance of both criteria, $t - t_d$ accounts for the number of processed frames since the descent onset, and $\bar{d}(t) = |D(t)|^{-1} \sum_{d \in D(t)} d$.

Throughout the descent, a permanent or momentary absence of the helipad in the visual field results in an instantaneous negative output of the detector based on the geometric moments and, in these situations, the UAV simply descends with a fixed latitude and longitude. When altitude a_l is reached, the decision is made as explained above.

VI. LANDING BEHAVIOUR

When landing the UAV, the helipad's structure overly occupies the entire field of view of the UAV's vision sensor, rendering ineffective its detection from the UAV's onboard camera. To provide 6-DOF accurate helipad-UAV relative pose estimates at very low altitude, a complementary strategy was devised. Namely, the landing platform itself assists the aerial vehicle in its final meters of the landing procedure.

A monocular camera with a wide field of view is placed at the helipad's centre to detect and track an augmented reality marker attached to the bottom centre of the UAV. The ArUco library [19], a third-party augmented reality toolkit, is used for the marker detection task. The pose estimate is sent to the UAV so that it is able to determine which corrective measures it must apply in order to centre itself on the helipad. Finally, when this marker is detected at the safe landing altitude, the UAV can safely turn off its rotors and land on the helipad.

To enable the use of a small marker, the landing behaviour just described is preceded by a preparatory move. Concretely, the UAV drops in altitude so as to be 2 m above the helipad, while maintaining its horizontal position from GPS-IMU feedback. The typical magnitude of this motion being small, around 2 m, potential horizontal displacements are robustly compensated by the wide field of view of the upwards looking camera.

Sudden moves outwards the camera's field of view resulting from strong winds need to be compensated. This is done by simply compensating the UAV's position from an EKF-based pose estimate and, when this does not suffice, the UAV is requested to climb in altitude to raise the chances of getting visible to the camera.

VII. EXPERIMENTS

This section presents a set of experiments carried out to validate the presented method in both simulated and real robots. Onboard the UAV, the system runs at 10Hz on a ODROID-XU (Quad core 1.6 Ghz Cortex-A7). The UAV tracking done by the helipad runs at 30Hz on a Quad core 3.5 Ghz i7. The system has been developed on top of the Robotics Operating System (ROS) and OpenCV. The input images are processed at a resolution of 640×480 . Vertical and horizontal fields of view of the upwards looking camera are 120° and 71° , respectively, and the diagonal field of view of the downwards looking camera is 69° .

The geo-referenced grid's dimensions are $(n, m) = (30, 30)$, with an area over ground in each cell of 900 m^2 . The helipad's area is 1 m^2 . The remaining parameters have been empirically defined as: $e = 1 \text{ m}$, $\sigma_v = 6$, $\beta = 0.5$, $a_h = 30 \text{ m}$, $a_m = 15 \text{ m}$, $a_l = 4 \text{ m}$, $\delta = 0.25$, $\xi = 7$, $\alpha = 0.3$, $\alpha_s = 0.7$, $\omega_v = 0.1$, $\omega_p = 0.47$, $\omega_l = 0.4$, $\omega_a = 0.03$, and $\tau = 0.2$.

A. Helipad detection and tracking

Two experiments were carried out in order test the ability of the presented helipad detection and tracking techniques. For the first experiment, the helipad was attached to a structure and laid down on open terrain. The terrain is populated with grassy patches, whose appearance is similar to the one of the helipad when seen from high above.

Fig. 2 depicts the system state in three moments of the takeoff sequence, one per row. The two top rows of the figure represent the phase of the takeoff in which learning is being guided by the detector used for close-in inspection. The red circles represent the output of the detector based on geometric moments (Section V-C). The third row refers to a latter moment during the takeoff, in which the detector based

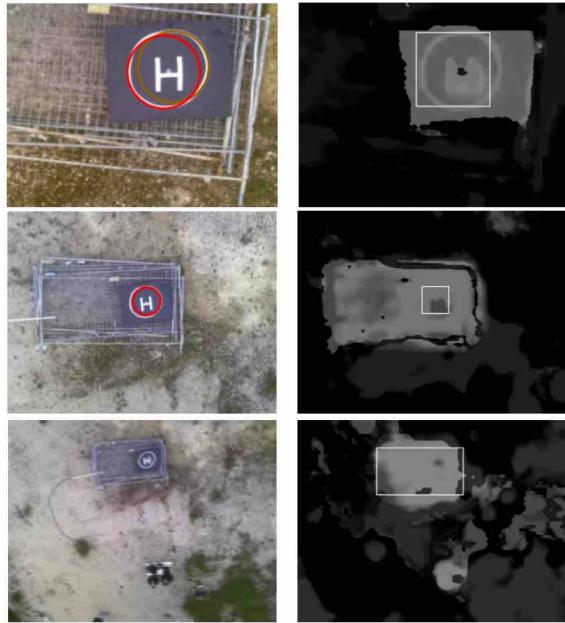


Fig. 2. Three key moments during takeoff (one per row). Left column: input image with helipad detection represented by the red circle overlay. Right column: saliency map of each input image with camshift tracker output represented by the overlaid rectangle. Saliency is represented by the brightness level of each pixel.

on geometric moments is no longer operational, as noticeable by the absence of the red circle overlay. In this case, learning is being guided by the camshift tracker deployed on the saliency map (Section V-C). Saliency maps are represented in the right column, and the overlaid rectangles represent the tracker's current state. Saliency being a regional descriptor, it is insensitive to considerable deformations observed in the input image (see the deformed H in the last row).

Then, the helipad was removed from its underlying structure and moved to a new location in the open field. Afterwards, the UAV was moved to this new location and asked to land on the helipad. The goal was to assess the robustness of the helipad detection techniques when the helipad's underlying and environment suffered slight changes. Fig. 3 depicts four key moments of the descent sequence. First, the UAV applies the altitude-dependent learned saliency computation weights and centres itself on the helipad according to the camshift tracker (first two rows). Once a mid-range altitude is reached, the detector based on geometric moments kicks in and controls the remainder of the descent (last row).

For the second experiment, the helipad was mounted on an ASV [5]. The experimental procedure was the same as for the first experiment. The UAV learned the helipad appearance while climbing in altitude (see Fig. 4) and used the result of the learning process to detect and track the helipad once this was covered by the UAV's field of view (see Fig. 5). At the maximum altitude, the helipad was mostly imperceptible to the UAV and still it was tracked successfully (first row of Fig. 5). This owes to the ability of the saliency learning process to capture the local context of the helipad, that is,

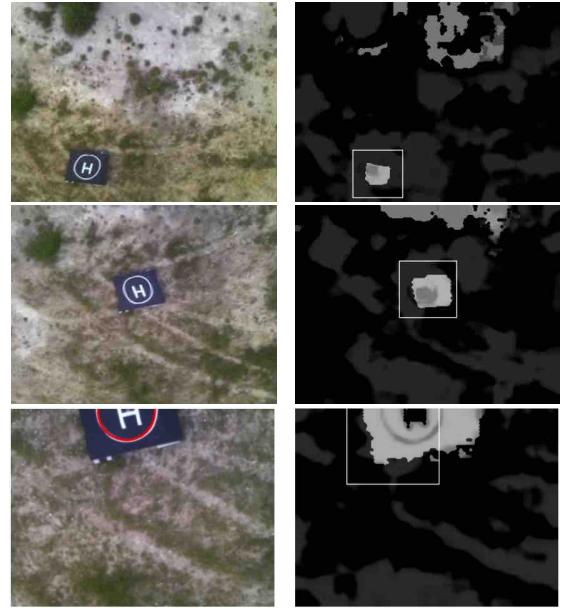


Fig. 3. Three key moments during landing (one per row).

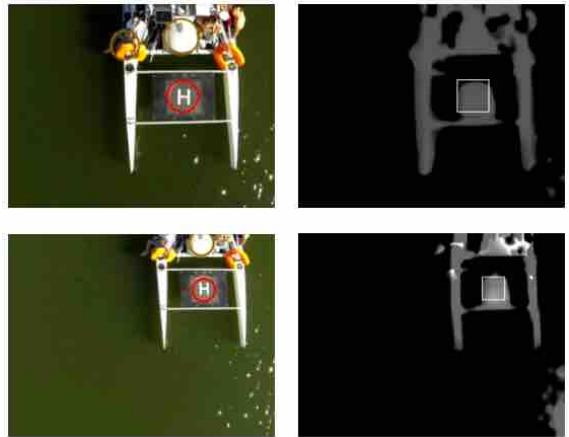


Fig. 4. Two key moments during takeoff (one per row).

the underlying mobile robot. It should be noted that only the appearance of the helipad was required at design.

Overall, the detector based on geometric moments successfully detected the helipad in 93 % of the frames it was applied across both experiments.

B. Assisted Landing

To assess the ability of the system to land the UAV based on the input images provided by the upwards looking camera mounted on the helipad, a set of landing trials were carried out. Fig. 6 depicts four key moments of a typical assisted landing.

Overall, the proposed method showed to be able to solve the landing problem. However, the experiments have revealed the need for an accurate control of the camera's exposure time and gain to properly cope with the fact that the marker is being observed from below.

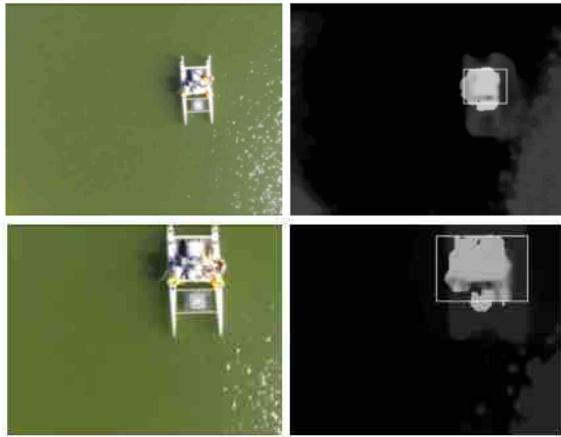


Fig. 5. Two key moments during descent (one per row).

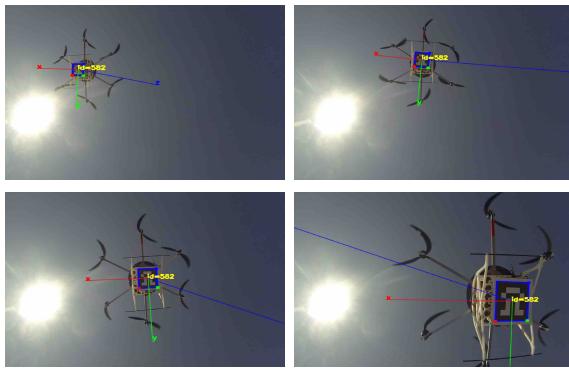


Fig. 6. Detection and tracking of the augmented reality marker in the last stage of the landing procedure. The referential frame overlaid represents the estimated pose. Time flows from left to right and from top to bottom.

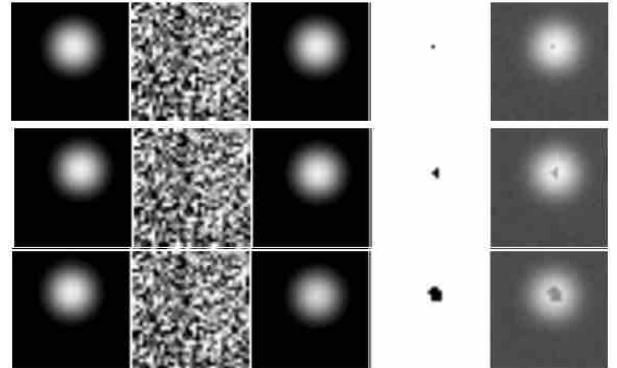
C. Global Search

A final experiment was run to assess the ability of the UAV to use the global search mechanism to find the helipad when outside its field of view (Section IV). For practical reasons, this experiment was run on the Kelpie simulator [20] (see Fig. 7). The simulated environment is composed of an ASV with a helipad on its top, another similar ASV but without a helipad, and a buoying cylindric object working as distractor.

In this experiment, the UAV was asked to perform a fully autonomous task. The task comprises taking off the helipad, moving towards a remote site, and then returning to the



Fig. 7. The simulated environment. Left: The ASV is the landing target and the floating cylinder a distractor. Right: A view of the helipad mounted on the ASV. The red circle overlay represents the output of the detector based on geometric moments.



$G_v(\cdot, \cdot, t)$ $G_a(\cdot, \cdot, t)$ $G_p(\cdot, \cdot, t)$ $G_l(\cdot, \cdot, t)$ $\Psi(\cdot, \cdot, t)$

Fig. 8. Three key moments in the global search (one per row). The first four columns depict the maps generated from the four criteria used to compute the global utility map present in the last column. The brightest pixel in the global utility map represents the location of the next waypoint for the UAV engage a local search. First row: global search onset, in which the next waypoint selected is slightly around the last known position of the helipad. The helipad is not found in this waypoint. Second row: intermediate moment in the global search, which exhibits a spiral-like diffusion. Third row: global search moment in which the helipad is found.

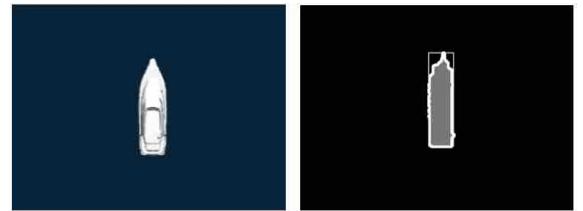


Fig. 9. A distractor attracting the UAV's attention during a local search. Left: input image captured by the UAV's downwards looking camera. Right: Object of interest segmented from the saliency map.

helipad and land on it. While the UAV flies away from the helipad, the ASV transporting the helipad was asked to move 70m away from its original position. As the last communication between helipad and UAV is assumed to have occurred at takeoff, the ASV's motion is unknown to the UAV.

As expected, the UAV returned to the last known position of the helipad, and locally searched for the helipad in a spiral-like pattern (see Fig. 8). The helipad being no longer on the same position, the local search failed to detect the helipad due to the presence of distractors (e.g., Fig. 9) or absence of any object of interest altogether. Consequently, the global search had to be reiterated in all these cases and a new vantage point selected for the next local search. Finally, the helipad was found on the ninth visited site. This shows that the global search procedure is capable of handling dynamic helipads.

Fig. 10 depicts a particularly interesting local search the UAV had to handle in the experiment. In this case the target ASV is nearby a distractor, thus both objects are labelled as of interest. However, the target exhibiting a higher average saliency level and bounding box similarity with the



Fig. 10. Local search with target and distractor in the UAV’s visual field.

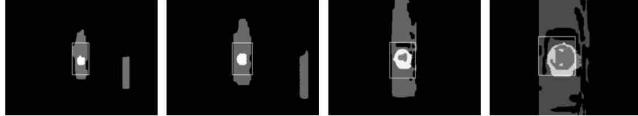


Fig. 11. Saliency maps during a close-in inspection. The overlaid rectangle represents the camshift tracker output. Time flows from left to right.

reference, results in gaining priority for the subsequent close-in inspection. As a result, the UAV centres itself on the ASV based on the camshift tracker deployed on the saliency map while simultaneously descending (see top row of Fig. 11). At the proper altitude the detector based on geometric moments detects the helipad and, so, it is able to guide the final phase of the descent. At the critical altitude a_l the UAV takes the final decision of landing.

VIII. CONCLUSIONS

A method for vision-based landing of a multirotor unmanned aerial vehicle on an autonomous surface vehicle (ASV) equipped with a helipad was presented and validated on both simulated and physical robots. To allow searching for the helipad when it is outside the UAV’s field of view, the method employs a global search procedure. This procedure fuses several criteria to pick the next best vantage point for performing a vision-based local search using the UAV’s downwards looking camera. Robustness to noise and image low resolution and deformation is attained by exploiting an saliency-based regional descriptor of the helipad, which is adapted by learning during the takeoff. Accurate pose estimation for the final landing phase is obtained by tracking the UAV from an upwards looking camera available at the helipad. As future work, we envision to: include the ability to search for a safe landing site; include local search from mosaics of aerial images; adapt the global search to consider the dynamics of the helipad; learn the method’s parameters from data; and perform extensive field trials.

ACKNOWLEDGMENTS

This work was co-funded by the ECHORD-RIVERWATCH FP7 project (grant nr. 231143) and the ROBOSAMPLER QREN project (LISBOA-01-0202-FEDER-024961). We would like to thank Eduardo Pinto for his contribution to the hardware development phase.

REFERENCES

- [1] J. Wendel, O. Meister, C. Schlaile, and G. F. Trommer, “An integrated gps/mems-imu navigation system for an autonomous helicopter,” *Aerospace Science and Technology*, vol. 10, no. 6, pp. 527–533, 2006.
- [2] G. Paravati, A. Sanna, B. Pralio, and F. Lamberti, “A genetic algorithm for target tracking in flir video sequences using intensity variation function,” *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 10, pp. 3457–3467, 2009.
- [3] T. Templeton, D. H. Shim, C. Geyer, and S. S. Sastry, “Autonomous vision-based landing and terrain mapping using an mpc-controlled unmanned rotorcraft,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2007, pp. 1349–1356.
- [4] A. Wendel, A. Irschara, and H. Bischof, “Natural landmark-based monocular localization for mavs,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2011, pp. 5792–5799.
- [5] E. Pinto, P. Santana, F. Marques, R. Mendonça, A. Lourenço, and J. Barata, “On the design of a robotic system composed of an unmanned surface vehicle and a piggybacked vtol,” in *Technological Innovation for Collective Awareness Systems*. Springer, 2014, pp. 193–200.
- [6] B. Herisse, T. Hamel, R. Mahony, and F.-X. Russotto, “Landing a vtol unmanned aerial vehicle on a moving platform using optical flow,” *IEEE Transactions on Robotics*, vol. 28, no. 1, pp. 77–89, 2012.
- [7] C. De Wagter and J. Mulder, “Towards vision-based uav situation awareness,” in *Proceedings of the AIAA Guidance, Navigation and Control Conference*, 2005, pp. 15–18.
- [8] Z. Yuan, Z. Gong, J. Wu, J. Chen, and J. Rao, “A real-time vision-based guided method for autonomous landing of a rotor-craft unmanned aerial vehicle,” in *Proc. of the IEEE Intl. Conf. Mechatronics and Automation*, vol. 4. IEEE, 2005, pp. 2212–2215.
- [9] S. Lange, N. Sunderhauf, and P. Protzel, “A vision based onboard approach for landing and position control of an autonomous multirotor uav in gps-denied environments,” in *Proceedings of the International Conference on Advanced Robotics (ICAR)*. IEEE, 2009, pp. 1–6.
- [10] S. Saripalli, J. F. Montgomery, and G. S. Sukhatme, “Visually guided landing of an unmanned aerial vehicle,” *IEEE Transactions on Robotics and Automation*, vol. 19, no. 3, pp. 371–380, 2003.
- [11] S. Saripalli and G. S. Sukhatme, “Landing a helicopter on a moving target,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2007, pp. 2030–2035.
- [12] L. Meier, P. Tanskanen, F. Fraundorfer, and M. Pollefeys, “Pixhawk: A system for autonomous flight using onboard computer vision,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2011, pp. 2992–2997.
- [13] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [14] P. Santana, L. Correia, R. Mendonça, N. Alves, and J. Barata, “Tracking natural trails with swarm-based visual saliency,” *Journal of Field Robotics*, vol. 30, no. 1, pp. 64–86, 2013.
- [15] S. Frintrop, G. Backer, and E. Rome, “Goal-directed search with a top-down modulated computational attention system,” in *Proceedings of the DAGM 2005, Lecture Notes on Computer Science, 3663*. Springer-Verlag, Berlin, Germany, 2005, pp. 117–124.
- [16] P. Santana, R. Mendonça, L. Correia, and J. Barata, “Neural-swarm visual saliency for path following,” *Applied Soft Computing*, vol. 13, no. 6, pp. 3021–3032, 2013.
- [17] J. G. Allen, R. Y. Xu, and J. S. Jin, “Object tracking using camshift algorithm and multiple quantized feature spaces,” in *Proceedings of the Pan-Sydney Area Workshop on Visual Information Processing*. Australian Computer Society, Inc., 2004, pp. 3–7.
- [18] S. Saripalli, J. F. Montgomery, and G. S. Sukhatme, “Vision-based autonomous landing of an unmanned aerial vehicle,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, vol. 3. IEEE, 2002, pp. 2799–2804.
- [19] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marin-Jimenez, “Automatic generation and detection of highly reliable fiducial markers under occlusion,” *Pattern Recognition*, vol. 47, no. 6, pp. 2280 – 2292, 2014.
- [20] R. Mendonça, P. Santana, F. Marques, A. Lourenço, J. Silva, and J. Barata, “Kelpie: A ros-based multi-robot simulator for water surface and aerial vehicles,” in *Proc. of the IEEE Intl. Conf. on Systems, Man, and Cybernetics (SMC)*. IEEE Press, 2013, pp. 1–8.