Letter to reviewers

Authors would like to thank the reviewers and editors for their efforts and the valuable help offered to improve our work. We consider that suggestions and comments have been of great relevance and the paper has been largely improved after the changes, corrections and additions. Authors are very grateful to the reviewers.

The changes in the paper have been highlighted in blue color.

Reviewer 1

ThiIn this paper the authors propose a monocular V-SLAM system with altitude measurements from a barometer and a vision-based distance measurement to a cooperative target. The landmarks are represented by both a Euclidean pose and inverse depth parameterization. The paper provides an observability analysis of the problem of the SLAM problem and a visual servo control law to follow the target. Finally, simulation and experiment data are provided. However, the strength of the contributions are weakened by the poor presentation and lack of rigour in the derivations.

1) ***The abstract, like much of the paper, can be improved by better defining the problem the paper is intending to solve. Phrases such as "the whole estimation problem" do not have any meaning.***

In the revised version of the paper, the abstract has been modified to better highlight the problem intended to be solved in the paper. Also, other modifications have been introduced along the text to improve clarity.

2) ***Furthermore, the actual state is only defined in the appendix (A2) and mentioned for the first time at the bottom of page 7.***

In order to improve the reading of the article, in the revised version of it, the definition of the system state (as well as the other content of the appendix) has been moved to Section “System specification”.

3) ***The literature review provides good references out of the vast literature on the subject, but downplays the success of other V-SLAM methods, and does not provide a strong motivation for why this exact setup is needed.***

In the revised version of the paper, a new paragraph has been added to the “Objectives and Contributions” section, which includes references to related V-SLAM methods in order to better highlight the contribution and motivation of the proposed approach.

4) ***The paper repeatedly mentions the importance of recovering the metric scale, but does not define the scale, nor directly show its estimation.***

The recovering of the metric scale is a very well-known problem for visual-based SLAM systems based on monocular cameras. And despite the vast literature about monocular-SLAM (many of them mention the scale problem), it is very surprising how very few of these works explicitly treat the scale problem. In this sense, the common approach is to compare the estimated trajectory with a reference trajectory in terms of the estimated position error. In this case, it is typically assumed that if the position error is small, then the metric scale has been somehow well recovered.

Although the position error and the recovery of the metric scale are related, the former cannot strictly be used as a direct measurement of the latter quality.

To address this previous issue, and to better highlight the performance of the proposed method, in the revised version of the paper, a new comparative study has been included. In this case, the proposed system is numerically compared to other related systems in terms of the convergence of the metric scale. For this purpose, the metric scale is explicitly computed.

Thanks to the reviewer for pointed out this matter. We considered this comment and the paper has been greatly improved.

5) ***The observability section is very difficult to read. At the very least it should start of by defining the state, the dynamics of the system, and the measurements. The majority of the variables are not defined.***

In concordance with this comment and also comment (2), in the revised version of the paper, the equations of the appendix have been moved to the main body, specifically to section “System specification”. Also, some descriptions of the formulae have been reworked. Moreover, the observability analysis section has been shortened, moving the descriptions of the Lie derivatives to an appendix. Now this section is more focused on the results.

We consider that with this change the readability of the observability section has been improved.

6) ***The observability in matrix O in (1) should be infinitely large. This is especially important when trying to show that something is not observable, and the higher order derivatives should only be dropped off with a proof showing that they do not increase the rank (also in line 147).***

We agree with the reviewer comment.

On the other hand, the analytical proof required to show that the rank of the observability matrix remains constant from some particular order derivative could be quite difficult given the complexity and size of these observability matrices.

Due to this fact, the rank of each matrix O was tested numerically in MATLAB. In this case, for each system configuration, starting with zero-order derivatives, the matrix O was being augmented each step using higher-order derivatives until its rank did not change. Moreover, the process was continued for several orders to ensure that some higher-order derivatives could not have contributed to the rank.

In the revised version of the paper, a clarification note has been added.

7) ***The rank calculations of O should be expanded upon (even if left to an appendix). It is important to point out at this point (3.2) that x and thus the dim(x) has never been defined.***

With no doubt, the thoroughness of the paper would benefit from the inclusion of the analytical proof of rank calculations. However, as we have mentioned in the previous answer, after conscientiously considering it, we believe that the proof would be excessively involving.

Also, it is important to note that we try to address the problem using a good balance of theory and practice, and therefore, theoretical, numerical, and experimental results are presented. In this sense, we consider that such analytical proofs fall out of the scope of the work, and thus the numerical validation should provide enough support.

8) Line 339 - "the right nullspace basis of O spans for N1" does not necessarily mean that all of the states are unobservable. The rank condition is only a sufficient condition of locally weakly observable.

We consider that this is a very interesting comment!

First, it is important to note that similar methodologies have been previously used in other related works to find the unobservable states of nonlinear systems. e.g.

- A. Cristofaro and A. Martinelli, "3D cooperative localization and mapping: Observability analysis," *Proceedings of the 2011 American Control Conference*, San Francisco, CA, 2011, pp. 1630-1635.
- Z. M. Kassas and T. E. Humphreys, "Observability Analysis of Collaborative Opportunistic Navigation With Pseudorange Measurements," in IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 1, pp. 260-273, Feb. 2014.

On the other hand, according to Hermann and Krener (1977), to prove that a system is locally weakly observable, it suffices to show that a submatrix of the observability matrix O, comprising a subset of its rows, is of full column rank. In contrast, to prove that a system is unobservable and find its unobservable directions, we need to: (a) Show that the infinitely many block rows of O can be written as a linear combination of a subset of its independent block rows, which form a submatrix O’, and; (b) find the nullspace of O’ to determine the system’s unobservable directions.

Although accomplishing (b) can be tractable, achieving (a) is extremely challenging especially for high-dimensional systems, as in the case of SLAM systems.

Therefore, contrary to the linear case, where it should only be derived no more than (n-1) times, due to Cayley Hamilton's Theorem. For nonlinear cases, unfortunately, this theorem doesn't exist, so it must compute the derivatives until reaching a full rank. To this effect, as it was mentioned before, due to the difficulty of analytically prove that the matrix O does not increase its rank after an n-order Lie derivative (i.e. prove (a)), a numerical test was carried out for finding the n-order derivative from which the rank does not increase.

Due to this fact, we are aware that in the absence of analytical proof of (a) we cannot rigorously prove the unobservable subspace.

On the other hand, we consider that the analysis presented in the paper still provides very useful qualitative information about the observability properties of the system. Moreover, the results obtained from the observability analysis are supported by the results obtained from the simulations.

In the revised version of the paper, an explicative note has been added to better clarify the limitations of the observability analysis. Thanks to the reviewer for bringing out this interesting comment!

9) ***Table 1 - shows that the state is position and velocity of the uav and target, but O appears to include all of the landmarks.***

The state of the system includes the landmarks, and in consequence matrix O also includes the landmarks. In table 1, the states related to the landmarks are those with a superscript “i”.

To improve the readability of the paper, and also in concordance with other related comments, the paper has been restructured. For instance, the system state, as well as the dynamical model and sensor models, has been now defined in section 2, before the observability section.

10)The observability study appear to conclude that the purpose of the work is measure the absolute position and velocity in the world, but this is unobservable. It would help the reader to point out that these values are not needed/important. The SLAM system is calculating the position and velocity within its own map. Also of importance is the relative distance between the target and the UAV, and this is partially measured. A stronger introduction/conclusion would really help the reader.

Section 3.2.3 (Remarks on the theoretical results) has been rewritten to better help the reader to interpret the observability results. Thanks to the reviewer for this suggestion.

11) ***For the sensors : Stating that the barometer can directly measure the height of the UAV (or that a sensor that directly measures the height even exists) is an assumption that should be explained. Barometers are similar to GPS's in that they both don't work indoors. Barometers also provide relative measurements and not absolute measurements.***

The strict requirement for the proposed method is the availability of altitude measurements respect to the reference system W (which is a local reference system). The absolute altitude measurements (e.g. respect to the sea level) can also be used if the absolute altitude of the reference system W is known.

Therefore, as a method, it is not strictly required that the altitude be obtained from a barometer. On the other hand, typically, altimeters are implemented using a barometer as its main sensor.

The raw output of a barometer is a signal proportional to the atmospheric pressure. Thus it is an absolute measurement. There is a relation between the absolute altitude and the atmospheric pressure, indeed.

The altimeter, as a device, processes the raw output of the barometer, by filtering it and compensating for errors like drifts due to temperature (e.g. using, for instance, temperature sensors) and producing an estimate of the altitude.

Altimeters provide measurements of absolute altitude (because barometers provide absolute pressure). But in practice altimeters can also provide local relative measurements of altitude, because they can be configured to use the local ground level (W reference) atmospheric pressure as their initial altitude reference *.

It is important to note that barometers can work well in common indoor environments because there is not such as a constant pressure atmosphere inside an indoor environment. On the other hand, barometers can be affected in periods of strong weather perturbations, such as the incoming of a storm when a variation in the atmospheric pressure occurs.

In the revised version of the paper, a note has been added to better explain the specification of the altitude measurements required by the method.

* Randal W. Beard and Timothy W. McLain, “Small unmanned aircraft; theory and practice” Princeton University Press, 2012.

12) ***This work also assumes a perfect non-drifiting AHRS would should be noted. Lastly, the distance to target measurement should explained along with its limitations.***

Typical available AHRS systems provide small-error and non-drifting attitude measurements for common applications. For these systems, the standard approach is to fuse highly accurate but long-term drifting gyroscopes measurements with noisy but almost non-drifting measurements obtained from accelerometers. The control system used for stabilizing UAVS, of multi-rotor type, uses the attitude computed by the AHRS system. If the AHRS system would provide drifting attitude measurements, it would be not possible to stabilize correctly the UAV.

Considering this fact, it is important to note that the proposed method assumes non-drifting but not perfect attitude measurements, which can be obtained from a standard AHRS system. In simulations, errors in the AHRS system as well as for the downward-looking camera assumption are emulated. The experiments with real data also validate the practical viability of the assumptions.

In the revised version of the paper, a couple of notes haven added to better clarify the AHRS assumption and limitations of the range measurements to the target.

13) ***Line:363 Add a cite for the "standard" EKF-based SLAM scheme.***

Done.

14) ***Line 176 - On the first read of the paper this was the first time I realized that the acceleration is considered as 0 mean noise. This basically implies that the UAV and target have 0 (or constant) velocity. This is a large limitation that should be discussed and implies large drift during dynamic manueuvers.***

The dynamical model used for the camera-robot (zero-mean acceleration noise) has become a standard model for this kind of application and it has been used in a lot of monocular SLAM approaches. This happens because the model is kept simple enough but still useful for many practical scenarios. In this case, since the model is used only to propagate the state and the covariance matrix through the prediction equations of the filter, the matrix Q (covariance of the process error noise) can be normally chosen to account for non-modeled dynamics. For instance, from our experience, for the performance of the SLAM system the measurement models are the most important part.

In case of the target model, we consider that a similar analysis applies, in the sense that this approach should be robust enough for several practical scenarios. It should be also considered that a cooperative target is assumed.

On the other hand, maybe a higher-order target model could be useful to better predict the trajectory of a highly maneuvering target, under periods where visual contact or range measurements could fail. On the other hand, in such an extreme scenario, it could be straightforward to augment the target model to consider higher-order dynamics.

In the revised version of the paper, a note about the former point has been added.

15) ***5 Target Tracking Control. A kinematic model is a poor choice of models for a quadrotor due to the coupling of its attitude and linear acceleration, but would also be appropriate for mostly hover flight. Why does (42) have mixed reference systems? psi dot = omega is an assumption and should be stated as such.***

A kinematic model was used because the proposed system control is a high-level control scheme that is used only for controlling the flight formation (generates the trajectory) of the UAV respect to the target.

In other words, it is assumed that a low-level control exists that effectively takes into account the coupling of the attitude and linear accelerations of the multi-rotor to controlling the linear velocities and yaw of the UAV given a reference input. In this case, the low-level control takes the control signal generated by the high-level control scheme as its reference signal.

In the revised version of the paper, a note has been added to better clarify the control strategy. Thanks to the reviewer for bringing out this potential confusion.

Results

16) ***There is no mention of velocity anywhere in the results, yet it is an important part of the state that is being estimated.***

In the revised version of the paper, a new Figure (figure 8) has been added where the velocity estimates and actual velocities are plotted for all the methods included in the comparative study.

17) Line 278: First mention of a gimbal and it's assumption. This is never explained.

Thanks to the reviewer for pointing out this error. The gimbal assumption was explained in the appendix but this did not make sense, indeed. So, to improve the readability, in the revised version of the manuscript, this assumption is now explained in section 2 (System specification) before being mentioned.

18) ***Line 288: If all of the initial conditions are exactly known then the initial positions of all of the landmarks in the camera should also be known exactly.***

Regarding this question, in simulations, the only assumption made is that the initial states of the camera-robot and the target are both known. Each time that the camera “sees” a new landmark, the system state is augmented with the initial landmark state computed as is explained in section 4.1. The initial landmark state is an initial estimate of the actual landmark location, with an associate initial error that depends on several sources of error. The filter minimizes the initial error when the landmark is re-observed in futures visual frames.

Only in the particular case of the test described in the subsection “Convergence and stability test”, the initial state of the landmarks ([x_0,y_0,z_0] position) is determined randomly, using a uniform distribution with different standard deviations. That means that landmarks are not initialized using any camera model, but instead are “artificially” added to the system state. In this case, the idea is to show how the robustness of the system takes advantage of the inclusion of the altimeter measurements, independently of the technique used for initializing the landmarks into the map.

In the revised version of the paper, the description of the experiment has been rewritten to better clarify this discussion.

19) ***Line 313: This needs to be more precise. The position is known exactly in the camera frame not the world frame.***

For our understanding, this comment is related with the previous one. to this effect, it is important to clarify that, as reviewer says, in visual SLAM the 3D position of the landmarks (in the world frame) is unknown a priori. On the other hand, the 2D projection (position) of the landmarks in the camera frame is not perfectly known because the visual measurement is affected in practice by errors like the optical distortion errors. In this case, it is important to note that in simulations these errors are emulated.

It is well known that for purely monocular SLAM systems (as the one used in the comparative study), the metric scale cannot be recovered without the use of some source of additional metric information. In this case, a classical approach is to initialize the monocular SLAM system with the states of perfectly-known landmarks called anchors. In this case, the term “perfectly-known” refers that the landmarks 3D position (in the world frame) is perfectly known a priori. New (unknown) landmarks added to the system state metrically converge respect to the anchors because the uncertainty of these anchors is

zero. In other words, the anchors have the effect of setting the metric scale of the estimates (both of the camera trajectory and the landmarks).

Our proposed method incorporates metric information into the system from the altimeter measurements as well as indirectly from the target measurements. And thus, it does not require anchors or any other kind of metric a priori information about the landmarks in order to recover the metric scale of estimates. Therefore, the anchors are used only in case of related methods, to make possible the comparison with the proposed method.

In the revised version of the paper, the description of the comparative study has been improved in order to better explain the methodology.

Also, some additional monocular-SLAM based methods that do not depend on anchors have been included now in the comparative study to better highlight the performance of the proposed method.

20) ***Line 318: This is an important point, but the theory and results in this paper do not show anything about scale. The converge properties of the scale is more nuanced then just motion along the vertical axis. In the comparative study is a little unfair to compare a strictly camera based method to a method that includes two other sensors. Other methods with altimeters and targets would be a better choice.***

We agree with the reviewer in the importance of including an explicit validation of the performance of the method for recovering the metric scale.

Since this topic was already discussed in the answer to comment (4), we will not add more information to the topic.

We consider that the paper has been greatly improved addressing the reviewer's comment.

21) ***Fig 8: Psi->There are some very large fast oscillations in the first second of flight. Is this physically possible or maybe dangerous for an actual quadrotor?***

This is an undesirable behavior for an actual UAV. Fortunately, the oscillation was just a plotting problem due to the misuse of the `atan` function instead of the `atan2` function, since the angle was jumping from a positive quadrant to a negative quadrant during the transient period.

Thanks to the reviewer for pointed out this issue.

22) ***Fig. 11: Z-> It appears to be more likely that the barometer was incorrect than the GPS. GPS problems are typically discrete jumps not slowly drifting up and down. It seems like this data shows a problem with the method (or at least the reliance on a barometer as an altimeter) rather than an advantage where GPS went wrong.***

The measurements obtained from the GPS on board the Bebop 2 are plotted correctly. In fact, it is well known that the vertical (z-axis) GPS error is larger than the horizontal axis errors*. Furthermore, the GPS measurements are following the expected response, since GPS errors are typically characterized as a combination of slowly varying biases and random noise**, or as a Gauss-Markov process***. We have also previously worked with GPS several times, and based in our experience, this is a normal GPS response.

Also, it is well known that most available autopilots make use of barometer-based altimeters (instead of GPS) for stabilizing the altitude of the multi-rotors when distance sensors (e.g sonar) are not available or are out of range.

Moreover, for experiments, the Bebop 2 (which has a sonar) was commanded to maintain a stable altitude during the trajectory. Accordingly, the altimeter measurements are in better concordance with this fact.

***** J. M. Zogg, GPS: Essentials of Satellite Navigation, u-blox AG, 2009.

** B.W. Parkinson et al. Global Positioning System: Theory and Applications, 1996.

***J. Ranking, “GPS and differential GPS: An error model for sensor simulation” in Position, Location and Navigation Symposium 1994.

23) ***Lastly, a large improvement to the paper would be a closed loop flight experiment using both the SLAM system and visual servoing.***

We agree with the reviewer's comment respect that a great addition to the paper would be a closed-loop flight experiment. In fact, that was our initial purpose for the experimental section of the paper. However, after successfully testing the SLAM system with our platform, we found troubles to access some elements of the driver of the platform required to implement the visual-servoing system. Currently, we are still trying to figure out how to solve this technical issue. And unfortunately, due to the short revision periods requested by the journal, it is not possible for us to accomplish the reviewer suggestion, neither by solving the current technical problem nor by using another experimental platform.

Actually, before submitting the manuscript to the journal, we considered two options: to remove the SLAM experiments with real data and present the paper with only simulations or remove the proposed control system and present a paper covering only the SLAM estimation technique. But finally, we decided that it is worth presenting all the available results (theoretical, numerical or experimental) for both the estimation and control systems.

A few minor issues:

- ***the paper needs heavy editing of English/grammar***
- ***after a formulas sentences are continued with "where" starting indented.***
- ***equation numbers squished down***
- ***"T" should not always be capitalized for target. Only when referring to "the" specific target defined***
- ***Font in the figures is not consistent with the text***
- ***Fig. 1: Equation numbers might help the readability. Also the notation is not that helpful as most variables are X. Text is not vertically aligned***
- ***Fig. 3: Different fonts, ZOH is not necessary.***
- ***Acronym style is not consistent.***
- ***make sure all notation is defined immediately after first use.***

In the revised version of the paper, most of the minor issues pointed out by the reviewer have been fixed. Due to the hard deadline for sending the revision and the time that had taken us to accomplish all the other reviewers' comments, some minor issues, as the font in the figures, have been left for the final version of the manuscript in case of acceptance.