Novel Hyperacute Gimbal Eye for Implementing Precise Hovering and Target Tracking on Board a Quadrotor

Augustin Manecy, Julien Diperi, Marc Boyron, Nicolas Marchand, Stéphane Viollet

HAL Id: hal-01261044

https://hal.univ-grenoble-alpes.fr/hal-01261044

Submitted on 23 May 2016

A Novel Hyperacute Gimbal Eye to Implement Precise Hovering and Target Tracking on a Quadrotor

Augustin Manecy[1,2,3], Julien Diperi[2], Marc Boyron[2], Nicolas Marchand[3], and Stéphane Viollet[2]

Abstract— **This paper presents a new minimalist bio-inspired artificial eye of only 24 pixels, able to locate accurately a target placed in its small field of view ($\pm 10°$). The eye is mounted on a very light custom-made gimbal system which makes the eye able to track faithfully a moving target. We have shown, that our gimbal eye can be embedded on a small quadrotor to achieve accurate hovering with respect to a target placed onto the ground. Our aiborne eye was enhanced with a bio-inspired reflex in charge of locking efficiently the robot's gaze onto a target and compensate for the robot's rotations and disturbances. The use of very few pixels allowed to implement a visual processing algorithm at a refresh rate of 400 Hz. This high refresh rate coupled to a very fast control of the eye's orientation allowed the robot to track a target moving at a speed up to $200°\cdot s^{-1}$.**

I. INTRODUCTION

Nowadays, many gimbal cameras with 2 or 3 degrees of freedom for Unmanned Aerial Vehicles (UAVs) are available on the market. Most of them provide a decoupling along the roll and the pitch angle and others provide also a degree of freedom to perform yaw movements. These systems are known under the name of *gimbal* or *pan-tilt system* and are intensively used for filming (sportive events, building inspection, etc.) These gimbal systems are often equipped with a dedicated GCU (Gimbal Control Unit) based on a dedicated Inertial Measurement Unit (IMU) strapped down on the camera and a dedicated controller, to make them autonomous and "easy-to-use" with different UAVs or UGVs. Once mounted on board an UAV and thanks to the additional IMU fixed to the camera mount, the camera gimbal is able, to compensate for all the rotations resulting from the UAV movements. The first application of these systems is the stabilization of camera for professional or amateur movies.

Many studies have been performed to achieve target tracking from a fixed wing UAV (for example [1], [2], [3]). [4], demonstrated that the position of a moving target (moving at $4-5 m \cdot s^{-1}$) can be estimated with a precision of 50m for a fixed wing UAV flying at an altitude of 400m. [5] worked on the estimation of the position of a target when the position provided by the the observer is uncertain (fixed wing UAV). They also proposed an evolution of the algorithm for a fixed wing UAV swarm ([6]). But gyro-stabilized camera are also used on rotorcraft, for example, [7] developed and validated in HIL a gimbal system for a helicopter UAV (based on the Yamaha R-max). Other studies involving quadrotors have shown that vision-based estimation can be combined with IMU measurements for state estimation ([8]) or track a moving person at a remarkable frame rate of 17 fps ([9]). Loianno *et al.* have shown that the very promising embedded software developed in the framework of the tango project could lead to an efficient stabilization of a quadrotor as well as creating 3D maps ([10]). Recent studies have proposed various vision-based control strategies to make a quadrotor able to follow a line ([11], [12]), to hover ([13], [14]), land automatically in a safe place ([15]) and track a moving robotic platform with active markers ([16], [17]). Unlike studies based on a fixed camera, we propose here a new approach where a novel visual sensor allows a small quadrotor to locate and track a non-emissive moving target with only 24 pixels. In addition, gaze stabilization of our gimbal eye does not need any additional strapped IMU. Quadorotor position estimation with respect to the target was achieved here by combining the visual error provided by the eye, the eye in robot orientation provided by the servomotors, the inertial measurements provided by the embedded IMU and an height measurement. Here, we present a new visual sensor and its innovative pan-tilt system which weight less than 50 g once assembled. This new eye is composed of: 1) a gimbal system, reproducing the decoupling between the head (center of vision) and the body existing on most animals, 2) a visual sensor inspired from the vision of the fly (for a review see [18]) and 3) a vibration mechanism which endows the visual sensor of hyperacuity, making it possible to locate accurately a target with only few pixels. The choice of using a gimbal system was motivated by the fact that it allows to separate the flight stability task (rejection of disturbances, etc.) from the target tracking task. We show that the eye was able to track the desired target, independently of the robot movements. It was demonstrated that this decoupling is probably one of the key points in bio-inspired stabilization strategies as shown by previous works on aerial robot ([19] and [20]).

Many multi-rotor robotic platforms performed tracking tasks by means of a fixed camera ([21], [22]), which can be tricky when it comes to track fast moving target. The use of a gimbal eye can overcome this issue and allow the visual tracking task to be independent of the robot displacements (see the very recent eXom drone developed by SenseFly which is equipped with a gimbal "head" [23]). This work is part of our steering by gazing strategy ([24], [25]) where the gaze orientation is directly used to solve the robot relative position with respect to the target, and let the possibility to the robot to realize a specific trajectory (turn around the target, stay at its vertical, stay at a certain distance, etc.)

Section II describes the motivations of using a bio-inspired

[1] Augustin Manecy, is with Onera - The French Aerospace Lab, F-31055 Toulouse, France, `augustin.manecy@onera.fr`

[2] Augustin Manecy, Julien Diperi, Marc Boyron and Stéphane Viollet, are with Aix-Marseille Université, Biorobotic Dept., CNRS, ISM UMR 7287, 13288, Marseille cedex 09, France, `<julien.diperi, marc.boyron, stephane.viollet>@univ-amu.fr`

[3] Augustin Manecy and Nicolas Marchand are with GIPSA-lab laboratory , Control Systems Dept., SySCo team, CNRS-Univ. of Grenoble, ENSE[3] BP 46, 38402 St Martin d'Hères Cedex, France, `nicolas.marchand@gipsa-lab.fr`

visual system and the related works. Then it describes the eye in details and the fusion algorithms used to locate the target. Then the section III describes the quadrotor used for the experiments, the control strategy and the gaze stabilization. Experimental results are given in section IV where our sighted aerial robot was seen to track autonomously a moving target.

Motivations and related-work

This work relies on a bio-inspired technology of visual sensor developed this last decade. This technology, named VODKA (for "Vibrating Optical Device for the Kontrol of Autonomous robots"), was introduced in [26] and [27] and some recent evolutions were described in [28] and [29]. These sensors feature an hyperacuity, which make them able to locate very accurately a target (an edge or a bar) with only one pair of pixels ([18]). The new visual system presented here, combines for the first time four elementary eyes, each composed of 6 pixels and a small lens (lens of the CMOS Camera Module SEN-00637), providing a total of 4×5 pairs of pixels (each eye is independant). Two eyes are disposed along one axis whereas the two other eyes are oriented along an orthogonal axis, making it possible to locate a target (e.g., a cross) in two different directions. Each "eye" tends to reproduce the main characteristics of the fly's optical properties (see [18]):

- each pixel has a Gaussian angular sensitivity,
- each pixel is submitted to the same retinal vibration responsible of an hyperacuity to locate contrasting objects (see review [18]),
- decoding of the visual information relies on pairs of adjacent pixels.

Using only 24 pixels drastically reduces the computational burden and thus, a refresh rate tested of 1 kHz can be obtained with a 16-bit microcontroller. A high refresh rate was essential for our robotic application where accurate hovering requires fast control loops and a precise estimation of the robot's position with respect to the target.

Compared to previous studies on hyperacute vibrating eyes ([26] or [28]), the main contribution of this work, concerns:

- a description of a robust calibration procedure, making the output signal provided by each eye perfectly linear with respect to the angular position of a contrasting bar (see section II-B),
- a new fusion algorithm to merge the different pairs of pixels and increase the field of view of the sensor (see section II-B),
- a new combination of four elementary eyes and an appropriate fusion (see section II-B), which extend the 1-D target localization to the 2-D case (azimuth-elevation).

II. A NEW GIMBAL VISUAL SYSTEM

A. Implementation of the gimbal Eye

The oculomotor system consists of an eye (the visual system) and a mechanical decoupling system to control the gaze orientation along two different axes (the roll orientation is denoted θ_{er_ϕ} and the pitch orientation is denoted θ_{er_θ}) independently from the robot's attitude. Figures 1a-c show the overall eye which consists of a ball joint and two spherical gimbals actuated by two drive-shafts connected to small servomotors of only 4 g (HiTec servo) with a rotation speed reaching up to $800° \cdot s^{-1}$ ($1000° \cdot s^{-1}$ without the eye) and a complete course of $\pm 50°$. To make the system lighter, the servomotors and the gimbals are directly fixed onto the electronic printed circuit board (PCB) of the eye (see figure 1c). The eye was fixed at the tip of a tube sliding along two semi-spherical gimbals. The other tip of the tube was fixed to a ball joint placed at the centre of the PCB. The mechanical coupling between the gimbals and the tube is ensured by a small brass ring sliding along the gimbals, and which also lock the rotation along the yaw axis. Finally, a small piston-spring system ensures a gentle contact between the brass ring and the gimbals in every angular position. A CAD view of the system illustrating the two degrees of freedom of the eye is given in figure 1a and a picture of the final eye prototype is given in figure 1c.

The eye consists of four independent linear pixel arrays (LSC iC-Haus) placed behind a lens (extracted from a CMOS Camera module SEN-00637). Two LSC are placed along one axis and two others along an axis orthogonal to the first one. The focal-length of each lens was adjusted to implement a Gaussian angular sensitivity $\Delta \rho$ of 3.7° for each pixel, and an inter-pixel angle $\Delta \phi$ of approximatively 3.9° (see figure 2). As described in (see [24], [28], [30]), we need an active vibration system
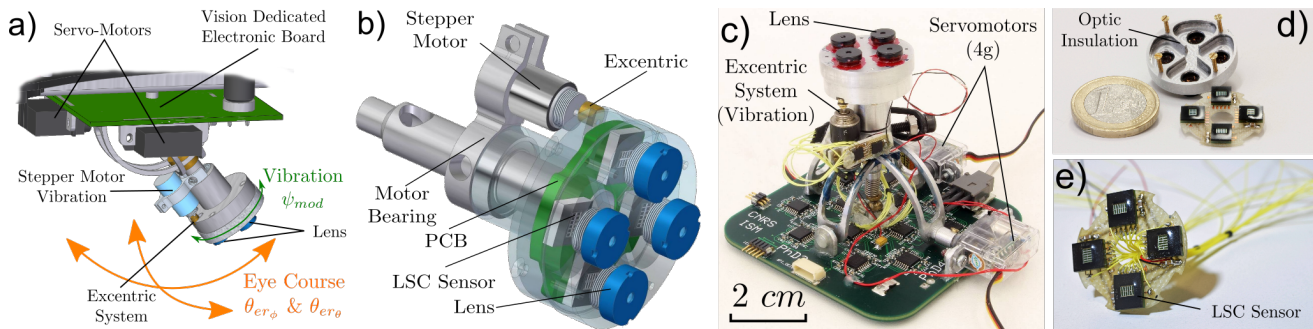

Fig. 1. a) CAD view of the complete airborne oculomotor system oriented toward the ground. The orientation of the 2-axis gimbal eye is controlled though two fast tiny servomotors (HiTec HS-5035HD). A stepper motor and an eccentric mechanism were used to impose to each pixel the same active vibration y. b) CAD view of the eye: the stepper motor combined with the eccentric mechanism yield a periodic (55 Hz) rotation of small amplitude (around 3.5°) making the lens of each eye translate periodically above the four linear arrays composed of 6 pixels (LSC sensor). c) The full airborne oculomotor system. d) Internal structure of the overall eye ensuring the optical insulation between the four elementary 1D pixel arrays. e) Small PCB supporting the four 1D arrays composed each of 6 pixels and its 26 wires (300 μm) connected to an external PCB supporting the visual processing unit.

endowing the eye with hyperacuity. The same actuator was used to generate simultaneously the vibration upon the 4 retinas. This strategy features several advantages:

- the vibration amplitude is strictly the same for the 4 eyes,
- the vibration is synchronous for the 4 eyes,
- the use of only one actuator allows to make the complete system more compact.
- the visual scanning frequency can be simply adjusted by changing the rotational speed of the stepper motor.

The micro-scanning of the four eyes is subject to active periodic rotational movements generated by a miniature eccentric mechanism. The angular vibration ψ_{mod} is generated by a tiny stepper motor (Faulhaber AM0820-V-5-56) with its driver (Allegro MicroSystems A3901) and an off-centered shaft, which translates along an elongated hole. The scanning frequency (55 Hz) can be easily adjusted by changing the rotational speed of the motor. The scanning amplitude depends on the diameter of the off-centered shaft.

B. Fusion and visual algorithms

1) ZSL function and visual signal fusion: The ZSL detector plays a key role in the fusion strategy of the LPU's output signals (Local Processing Unit). Originally described by [31] and [24], this non-linear function was modified to be adapted to the multi-LPU case. In this study, the ZSL function plays two different roles:

- Determine the best LPU to be used for the localization algorithm.
- Detect if a target is present or not in the field of view and set the signal to zero if there is no target (see [32]).

The implementation of the ZSL results from the fact that for each LPU the sum of the demodulated pixels output signals i and j (denoted $\Sigma Ph_{i,j}$) is maximal (with respect to the other LPU) when a contrasting edge or bar is placed in its local FOV. As a consequence, the LPU used for the localization is selected by looking for the maximum of the sum among the 5 LPUs for each eye (i.e., maximum of $\Sigma Ph_{i,j}$). Figure 3a shows the sum of each LPU with respect to the angular position of the target. The ZSL output signal corresponds to the index of the selected LPU as shown in figure 3b. But the selected LPU is used only if one of the adjacent LPU has also a sum higher than a predefined value. This simple method allows eliminating false detections of the target.

2) Calibration procedure: The output signal of a LPU is a nonlinear function depending on the angular position of the target, and on the optical parameters $\Delta\rho$ and $\Delta\phi$ (see [26]). As a consequence, the output of each LPU is linearized by using a look-up table corresponding to a calibration table. These look-up tables are built during a calibration phase, which is performed for a given distance between the eye and the target. The calibration procedure consists of making the eye rotate slowly in front of a target at a known distance. During this phase, the robot is strictly immobile, and the angular position of the eye is then equivalent to the angular position of the target in the eye's frame. The look-up table is built by recording both about 100 angular positions of the eye and the corresponding filtered LPU output signals. As each LPU has different optical parameters, each LPU has its own calibration table. Figure 3c shows the output signal provided by the eye noted 1 before and after the calibration.

Once the calibration was performed, the eye's output signal follows faithfully the target angular position. It is worth noting that the output signal is set to 0 when the target is out of the field of view thanks to the ZSL.
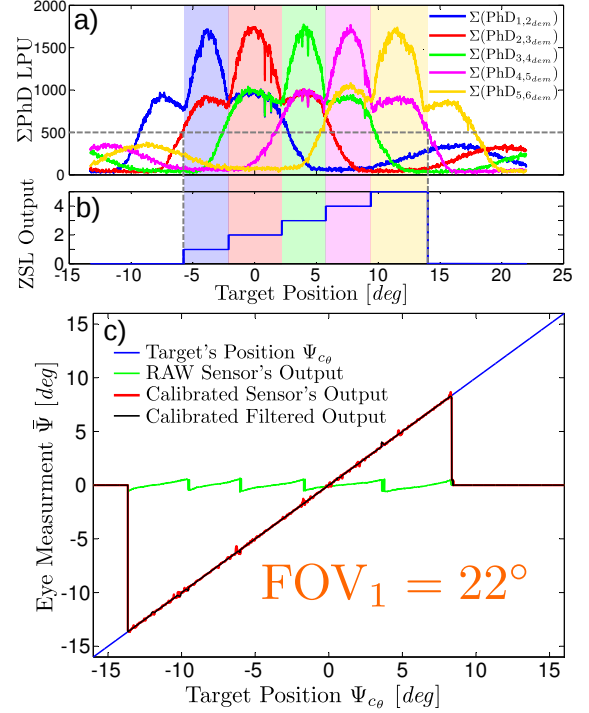

Fig. 3. a) The sum of demodulated pixels output signals for each LPU. The selected pair of pixels corresponds to a pair with the maximum sum and at least one of its adjacent pair with a sum higher than a predefined threshold (gray dotted line). b) ZSL output signal, corresponding to the index of the selected LPU depending of the angular position of the target. c) The measured angular position of one elementary eye versus the actual target position. The actual angular position of the target (blue), the measured eye's output signal (green), the calibrated output signal processed through the look-up table (red), and the filtered calibrated output signal (black). The ZSL sets the eye's output signal to zero when the target is out of the field of view.

3) Multi-eye fusion: The visual system is composed of 4 eyes, which measure the orientation of a target according to 2 directions. The eyes 1 and 3 measure its orientation along the pitch axis (θ_{er_θ}) and the eyes 2 and 4 along the roll axis (θ_{er_ϕ}). Then, these 4 measurements can be fused to improve the final target orientation measurement. We noted $\bar{\epsilon}_{r_i}$ the angular position of a contrasting bar placed in the field of view of the eye i, then the orientation of the target along the roll and the pitch axis can be obtained as follows:

$$\bar{\phi} = \frac{1}{2}(\bar{\epsilon}_{r_2} + \bar{\epsilon}_{r_4}) \tag{1}$$

$$\bar{\theta} = \frac{1}{2}(\bar{\epsilon}_{r_1} + \bar{\epsilon}_{r_3}) \tag{2}$$

The eye is able to locate a target with an accuracy of 0.1°, which is 39 times smaller than the inter-pixel angle, at a refresh rate as high as 1kHz. However, the state estimation of the robot and the feedback loops used to stabilized its attitude were running onboard the quadrotor at a sampling frequency of 400 Hz, so the refresh rate of the visual processing was limited here at 400Hz.
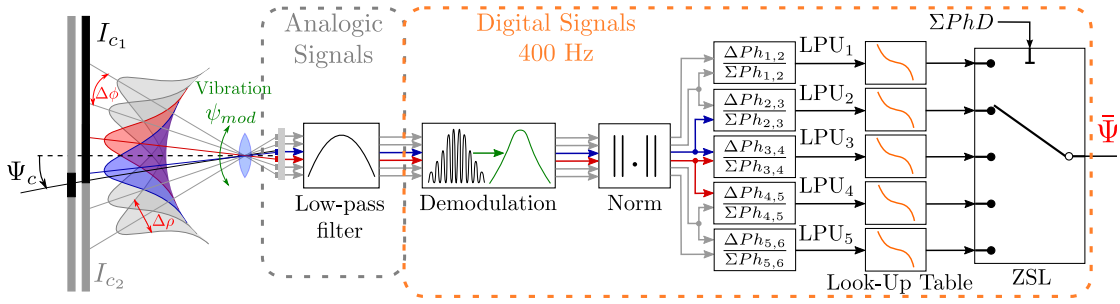
Fig. 2. Block diagram representing the visual processing of the different pairs of pixels, called LPU (Local Processing Unit). The output signals of each pixel is demodulated and normalized. Then, for each pair, the difference $\Delta Ph_{i,j}$ divided by the sum $\Sigma Ph_{i,j}$ of adjacent pixels is computed. The angular position of a target measured by an LPU is obtained by using a calibration look-up table (computed during a calibration phase). Finally, the ZSL function selects the best pair, or set the output $\bar{\Psi}_c$ to 0 if there is no target in the visual field.

III. THE QUADROTOR X4-MAG

A. Description of the complete system

The quadrotor used in this work is an opensource aerial robotic platform, called X4-MaG, described in details in [33] and its autopilot is fully developped with MATLAB/Simulink thanks to the open-source RT-MaG toolbox ([34]). We mounted the eye described in section II on board the robot X4-MaG. The eye was directly connected to the high level controller of the robot (the Gumstix Overo AirSTORM) via an SPI link running at 12 MHz to process the visual output signals provided by the eye.
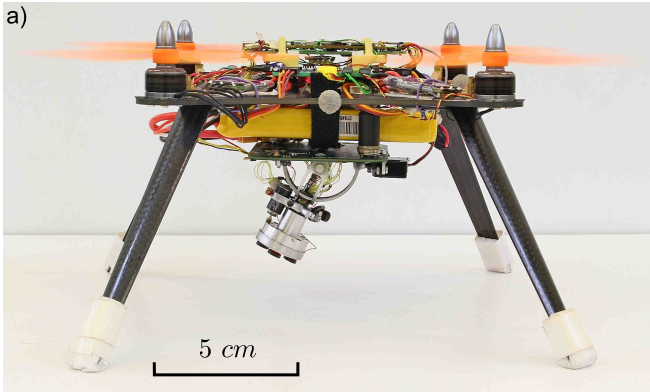

Fig. 4. The X4-MaG quadrotor and its bio-inspired eye weight around 395 g, has a 30-cm span and an autonomy of about 10 min. The robot is equipped with an Arduino board (NanoWii), using the MPU6050 IMU, as low-level controller and a Gumstix Overo as high-level controller. Its eye of 24 pixels localizes a contrasting cross with an accuracy of 0.1° and is able to track it in to directions for more than 100°. The robot has a manual pilot mode, an autopilot mode based on a motion capture system (VICON), and an autonomous mode based on its IMU and its vision system.

B. Control Strategy and state estimation

1) Attitude and position controller: A quaternion-based complementary filter was used to fuse the measurements of the 6-axis IMU and the yaw measurement of the VICON system (we used the yaw measurement provided by the ground-truth VICON system instead of a 3-axis magnetometer not available onboard). A geometric controller based on quaternions performs the attitude control loop. The position control loop consists in a saturated PID controller yielding acceleration setpoints which

are then converted to attitude setpoints by inverting the nonlinear model of the quadrotor. The whole control strategy and attitude estimation is described in details in [33]. Our novel control strategy takes benefit from the decoupled eye to estimate the robot's position and linear speed. This strategy is described in section III-B.2

2) Position estimation: Assuming that the distance to the target is known and measured by a motion capture system with a high accuracy, the position and the velocity of the robot are estimated by using the quadrotor dynamic model and the measurements (retinal errors $\epsilon_{r\phi}$ and $\epsilon_{r\theta}$ and orientations $\theta_{er\theta}$ and $\theta_{er\theta}$) provided by the eye. The retinal errors $\epsilon_{r\phi}$ and $\epsilon_{r\theta}$ are defined as the angular position of the target in the eye frame. It also can be seen as the angular error between the target position and the gaze direction, as defined in [25].

The model of the quadrotor is given by the following equations:

$$
\begin{pmatrix} \dot{\hat{X}} \\ \dot{\hat{Y}} \\ \dot{\hat{Z}} \\ \dot{\hat{V}}_X \\ \dot{\hat{V}}_Y \\ \dot{\hat{V}}_Z \end{pmatrix} = \begin{pmatrix} \hat{V}_X \\ \hat{V}_Y \\ \hat{V}_Z \\ \frac{\hat{T}}{m}(\sin\hat{\phi}\sin\bar{\psi} + \cos\hat{\phi}\cos\bar{\psi}\sin\hat{\theta}) - \frac{K_x}{m}V_X \\ \frac{\hat{T}}{m}(\cos\hat{\phi}\sin\bar{\psi}\sin\hat{\theta} - \sin\hat{\phi}\cos\bar{\psi}) - \frac{K_y}{m}V_Y \\ \frac{\hat{T}}{m}(\cos\hat{\phi}\cos\hat{\theta}) - \frac{K_z}{m}V_Z \end{pmatrix} \quad (3)
$$

The estimated state vector $\hat{x}$ is given by $\hat{x} = \begin{pmatrix} \hat{X} & \hat{Y} & \hat{Z} & \hat{V}_X & \hat{V}_Y & \hat{V}_Z \end{pmatrix}^T$. $\hat{T}$ is the estimated thrust corresponding to a given rotor speed, m is the mass of the robot, and K_x, K_y and K_z are the equivalent drag coefficients resulting principally from the flapping angles of the propellers.

The actual position of the servomotors θ_{er_ϕ} and θ_{er_θ} corresponding to the gaze orientation can be estimated by using the orientation setpoints ($\theta^\star_{er_\phi}$ and $\theta^\star_{er_\theta}$) and the dynamic model of the servomotors. Then, we can merge these orientations, the altitude $\bar{Z}_{\text{Vicon}}$, the estimated attitude and the measured retinal errors ($\bar{\epsilon}_{r_\phi}$ and $\bar{\epsilon}_{r_\theta}$) to build pseudo measurements of the current positions assuming that the ground height is known:

$$
\bar{y} = \begin{pmatrix} \bar{X} \\ \bar{Y} \\ \bar{Z} \end{pmatrix} = \begin{pmatrix} \bar{Z}_{\text{Vicon}}.\tan(\hat{\theta}_{er_\theta} + \bar{\epsilon}_{r_\theta} + \hat{\theta}) \\ -\bar{Z}_{\text{Vicon}}.\tan(-\hat{\theta}_{er_\phi} - \bar{\epsilon}_{r_\phi} + \hat{\phi}) \\ \bar{Z}_{\text{Vicon}} \end{pmatrix} \quad (4)
$$

The state observation is then performed by a standard Extended Kalman Filter (EKF) based on the model described by (3) and (4).

C. Gaze stabilization and visual tracking

The robot's gaze was controlled by merging two different reflexes. A gaze stabilization reflex, which was in charge to compensate for all the robot rotations, and a visual tracking reflex to make the eye locked onto a moving target.

1) Gaze stabilization: We assumed that the center of gravity of the robot was close to the center of rotation of the eye. As a consequence, the gaze stabilization reflex (vestibulo-occular reflex, denoted VOR) is implemented as an orientation setpoint of the eye equal to the opposite of the current robot's attitude. Therefore, the reflex is implemented for each axis as a feedforward control:

$$\theta^\star_{er\phi\_\text{VOR}}(t) = -\hat{\phi}(t) \tag{5}$$

$$\theta^\star_{er\theta\_\text{VOR}}(t) = \hat{\theta}(t) \tag{6}$$

Where $\hat{\phi}$ and $\hat{\theta}$ are the estimated roll angle and the estimated pitch angle in the inertial frame $\mathcal{I}$, and $\theta_{er\phi\_\text{VOR}}$ and $\theta_{er\theta\_\text{VOR}}$ are the setpoints in the fixed eye frame $\mathcal{F}_e$.
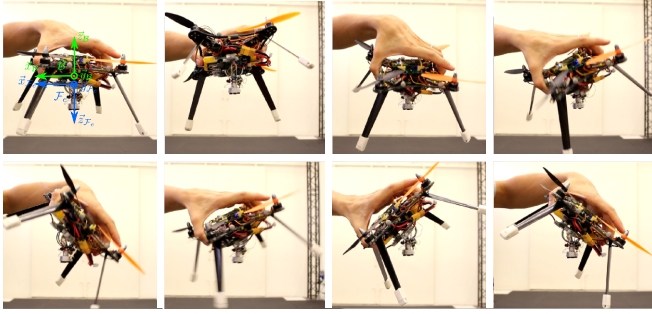
Fig. 5. The gaze stabilization acts so to hold the gaze along its absolute orientation even if the robot rotates in pitch and roll (here the gaze was maintained vertically). Unlike a classical gimbal camera, the gaze stabilization is achieved here without any additional IMU strapped onto the eye.

2) Visual tracking: A visual fixation reflex denoted VFR, makes the robot's eye locked onto a target and compensate for any robot or target translations. The visual tracking task implements a visual feedback loop which keeps the retinal errors $\epsilon_{r\phi}$ and $\epsilon_{r\theta}$ close to 0 (that is mean this control keeps the target in the center of the eye's field of view). The two axes (roll and pitch) are assumed to be independent, which means that there is an independent saturated PI controller for each axis:

$$\theta^\star_{er\phi\_\text{VFR}}(t) = k_p \cdot \left(\tilde{\epsilon}_{r\phi}(t) + k_i \int_0^t \tilde{\epsilon}_{r\phi}(t)dt \right) \tag{7}$$

$$\theta^\star_{er\theta\_\text{VFR}}(t) = k_p \cdot \left(\tilde{\epsilon}_{r\theta}(t) + k_i \int_0^t \tilde{\epsilon}_{r\theta}(t)dt \right) \tag{8}$$

Where $\tilde{\epsilon}_{r\phi}(t) = \epsilon^\star_{r\phi}(t) - \bar{\epsilon}_{r\phi}(t)$ and $\tilde{\epsilon}_{r\theta}(t) = \epsilon^\star_{r\theta}(t) - \bar{\epsilon}_{r\theta}(t)$ with $\epsilon^\star_{r\phi}(t) = \epsilon^\star_{r\theta}(t) = 0 \; \forall t$. The final orientation setpoint for each axis was then obtained by combining the gaze stabilization reflex (VOR) and the VFR:

$$\begin{cases} \theta^\star_{er\phi}(t) = \theta^\star_{er\phi\_\text{VOR}}(t) + \theta^\star_{er\phi\_\text{VFR}}(t) \\ \theta^\star_{er\theta}(t) = \theta^\star_{er\theta\_\text{VOR}}(t) + \theta^\star_{er\theta\_\text{VFR}}(t) \end{cases} \tag{9}$$

Due to limitations of the servomotors, this closed-loop control of the eye was performed at 333 Hz even if the visual localization of the target was performed at 400 Hz. Finally the eye was able to locate a target with a 0.1-° precision along two directions over 100° (thanks to its decoupling). The eye was also able to track a target moving up to $200° \cdot \text{s}^{-1}$.

IV. APPLICATION TO TARGET TRACKING

A. Experimental conditions

The oculomotor system (the eye) was attached under the robot by means of three tiny snap magnets. Figure 4 shows the robot with its eye looking downward. The complete robot weighs approximately 395 g for a span of 30 cm and an height of 17 cm. With this configuration, its flight autonomy is about 8-10 min.
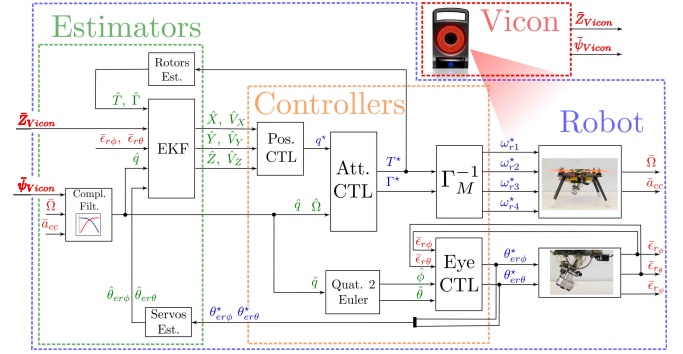
Fig. 6. Block Diagram of the complete system. The `Compl. Filt.` block estimates the robot's attitude, `Servos Est.` estimates the actual servo position based on their dynamic model; and the `Rotors Est.` estimates the thrust and torques based on the dynamics of the rotors. The `EKF` is an extended Kalman filter which estimates the position and the translational speed of the robot. `Pos. CTL` is the position controller, `Att. CTL` the attitude controller and `Eye CTL` the gaze controller. `Quat. 2 Euler` extracts the corresponding Euler angles from the estimated attitude quaternion and Γ_M^{-1} converts the torques and thrust setpoints into rotational speed setpoints for the four rotors. The VICON system is used here as a ground truth system, but also to deliver the yaw angle and the altitude of the robot.

B. Positioning and Tracking Results

Assuming the robot height is provided by the motion capture system (VICON) of the flying arena, the robot is able to estimate its position with respect to the target by using only its eye and its unique embedded IMU strapped onto the robot's body. Using the height measurement of the VICON system may seem very restrictive but new generation of distance sensors (as [35]) can be used to provide a reliable height measurement, event for aggressive manoeuvers [36]. To evaluate the performances of the estimation and tracking algorithms, the motion capture system was also used as a ground truth to monitor the position of the target and the position of the robot.

Scenario: The robot was placed above the target and took off, then reached a height of 1.3 m above the target. Then, at $t = 245$ s, it performed a 1-m step with respect to the target along the Y axis, and came back after 45 s. Finally, the target moved along the X and Y axis in order to evaluate the tracking performances of the quadrotor. Then the robot landed automatically.

Figure 7 shows the evolution of the robot's and target's positions during the scenario composed of three main steps.
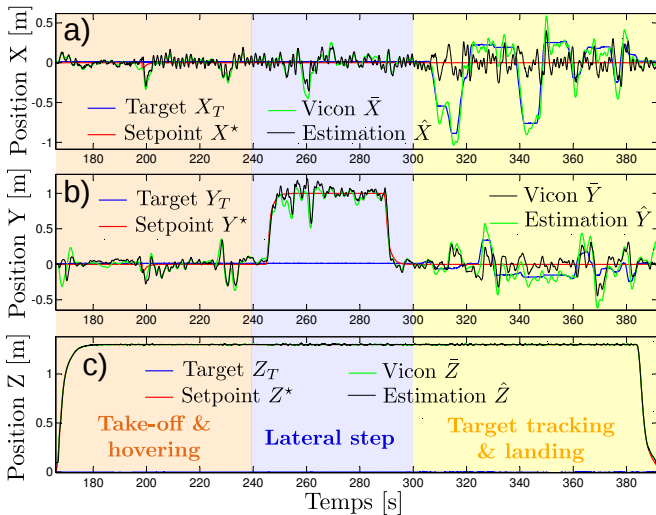
Fig. 7. a) Time course of the robot and target positions along the X axis, from $t = 300s$ to the end. The robot keeps its estimated position relative to the target close to 0 (black curve) and follows the moving target faithfully (blue curve). b) Time course of the different positions along the Y axis. c) Time course of the positions along the Z axis.

a)	$\hat{\theta}_{er_\phi}$	$\hat{\theta}_{er_\theta}$	$\bar{\epsilon}_{r_\phi}$	$\bar{\epsilon}_{r_\theta}$
Mean	9.54°	-0.29°	-0.0027°	-0.0084°
STD	17.14°	7.06°	1.31°	1.65°
Min	-27.41°	-31.21°	-5.61°	-8.12°
Max	50.10°	31.88°	5.90°	8.63°

b)	$\tilde{X}$ [cm]	$\tilde{Y}$ [cm]	$\tilde{V}_X$ [cm·s^{-1}]	$\tilde{V}_Y$ [cm·s^{-1}]
STD	4.41	5.89	9.78	9.97
Min	-23.27	-30.78	-37.87	-33.69
Max	12.00	16.28	40.09	8.63

TABLE I

Performances of the different estimations and control loops. With $\tilde{X} = \bar{X} - (\hat{X} + \bar{X}_T)$ and $\tilde{Y} = \bar{Y} - (\hat{Y} + \bar{Y}_T)$ and $\tilde{V}_X = \bar{V}_X - (\hat{V}_X + \bar{V}_{X_T})$ and $\tilde{V}_Y = \bar{V}_Y - (\hat{V}_Y + \bar{V}_{Y_T})$.

Thanks to the VICON system, the altitude of the robot ($\bar{Z}$) was controlled in closed-loop with a high accuracy (see figure 7c). During the take-off and the hovering phase, we observed that the positions estimated by the robot were close to the ground-truth positions measured by the Vicon system. As shown in the blue area of the figure 7b, the robot was also able to move around the target automatically. During the target tracking phase, the robot kept its position relative to the target close to zero (black curves in the yellow part of figure 7), which means that the robot is flying above the target (i.e., the robot follows the target). This is confirmed by the ground truth of the robot position (green bine) which follows the ground truth of the target position (blue line). Figures 7a and 7c show that the green curve follows the blue one faithfully.

Table I describes the mean retinal errors ($\bar{\epsilon}_{r_\phi}$ and $\bar{\epsilon}_{r_\phi}$) and the estimation errors I. During flight, the eye tracks faithfully the target as shown by the standard deviation of only 2°, and the maximum error is below 9°. The position estimation errors feature a good precision with a standard deviation as small as 6 cm. During the 1-m lateral translation, the robot makes a stiff roll angle leading an orientation of the eye along the roll axis (θ_{er_ϕ}) greater than 50° achieved within less than 0.6 s. But during this fast and large rotational movement (more than 80°·s^{-1}), the retinal error ϵ_{r_ϕ} was kept lower than 6° demonstrating the high efficiency of the gaze control system and the mechatronics of the eye.

C. Discussion

For the sake of clarity and simplicity, we did not deal with the retinal error of the yaw in this paper (yaw misalignement between target and robot). But this error can be measured by combining the four retinal errors measurments of the different eyes and using some characteristic distance of the eye's geometry. Then, to compensate for yaw movements of the target, the robot has to adjust its own yaw angle, to maintain the yaw retinal error close to zero. As there is no additional yaw decoupling in

the gimbal systemand the field of view of the eye is about only $\pm20°$ with respect to the yaw axis, only small yaw disturbances can be rejected and only slow yaw motion of the target can be tracked. Adding a yaw decoupling would allow the robot to reject much larger yaw disturbances and track a rotation of the target around the vertical axis.

It was shown in this paper that a simple target (a white cross above a black background) can be tracked efficiently using a quadrotor platform and its minimalist gimbal eye system. Thanks to the analogic pass-band filter the system is also robust to illumination change. However, to be able to track more complex targets, in natural environments, with texture and and objects, this simple eye could be replace by another minimalistic bio-inspired eye [37]. Indeed, it was proven that such an eye was able to perform visual odometry tasks, and could also be used to track a target moving over a textured background [30].

V. CONCLUSION

In this paper, we presented a complete solution to track with a quadrotor a known target moving on the ground. A bio-inspired eye of only 24 pixels mimicking some features of the fly's vision performs the visual tracking task. Using only few pixels and fast algorithms make it possible to locate a target at a high refresh rate (here at 400 Hz) which drastically improves the performances of the visual tracking. This eye was mounted on board a quadrotor, and considered as an independent system used to track the target thanks to its mechanical active decoupling (a gimbal system). The estimation of the attitude is done classically by merging the data of only one IMU, and the position of the robot (relative to the target) is estimated thanks to the gaze orientation. We demonstrated experimentally that this system can be used efficiently to track a moving target and is also able to land accurately on this target (see attached video).

REFERENCES

[1] I. Wang, V. Dobrokhodov, I. Kaminer, and K. Jones, "On vision-based target tracking and range estimation for small uavs," in *Guidance, Navigation, and Control and Co-located Conferences*. American Institute of Aeronautics and Astronautics, Aug. 2005, pp. –.

[2] F. Rafi, S. Khan, K. Shafiq, and M. Shah, "Autonomous target following by unmanned aerial vehicles," *Proc. SPIE*, vol. 6230, pp. 623 010–623 010–8, 2006.

[3] M. Quigley, M. Goodrich, S. Griffiths, A. Eldredge, and R. Beard, "Target acquisition, localization, and surveillance using a fixed-wing mini-uav and gimbaled camera," in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, April 2005, pp. 2600–2605.

[4] V. Dobrokhodov, I. Kaminer, K. Jones, and R. Ghabcheloo, "Vision-based tracking and motion estimation for moving targets using small uavs," in *American Control Conference, 2006*, June 2006, pp. 6 pp.–.

[5] M. Campbell and M. Wheeler, "A vision based geolocation tracking system for uav's," in *Guidance, Navigation, and Control and Co-located Conferences*. American Institute of Aeronautics and Astronautics, Aug. 2006, pp. –.

[6] M. Campbell and W. Whitacre, "Cooperative tracking using vision measurements on seascan uavs," *Control Systems Technology, IEEE Transactions on*, vol. 15, no. 4, pp. 613–626, July 2007.

[7] O. Jakobsen and E. Johnson, "Control architecture for a uav-mounted pan/tilt/roll camera gimbal," in *Infotech@Aerospace Conferences*. American Institute of Aeronautics and Astronautics, Sep. 2005, pp. –.

[8] J. Andersh, A. Cherian, B. Mettler, and N. Papanikolopoulos, "A vision based ensemble approach to velocity estimation for miniature rotorcraft," *Autonomous Robots*, vol. 39, no. 2, pp. 123–138, 2015.

[9] H. Lim and S. N. Sinha, "Monocular localization of a moving person onboard a quadrotor mav," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, May 2015, pp. 2182–2189.

[10] G. Loianno, G. Cross, C. Qu, Y. Mulgaonkar, J. Hesch, and V. Kumar, "Flying smartphones: Automated flight enabled by consumer electronics," *Robotics Automation Magazine, IEEE*, vol. 22, no. 2, pp. 24–32, June 2015.

[11] I. Sa and P. Corke, "100hz onboard vision for quadrotor state estimation," in *Australasian Conference on Robotics and Automation*, 2012.

[12] L. Carrillo, G. Flores, G. Sanahuja, and R. Lozano, "Quad-rotor switching control: An application for the task of path following," in *American Control Conference (ACC), 2012*, June 2012, pp. 4637–4642.

[13] S. Azrad, F. Kendoul, D. Perbrianti, and K. Nonami, "Visual servoing of an autonomous micro air vehicle for ground object tracking," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, Oct 2009, pp. 5321–5326.

[14] J.-E. Gomez-Balderas, S. Salazar, J. A. Guerrero, and R. Lozano, "Vision-based autonomous hovering for a miniature quad-rotor," *Robotica*, pp. 43–61, Jan. 2014.

[15] C. Forster, M. Faessler, F. Fontana, M. Werlberger, and D. Scaramuzza, "Continuous on-board monocular-vision-based elevation mapping applied to autonomous landing of micro aerial vehicles," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, May 2015, pp. 111–118.

[16] J. Daly, Y. Ma, and S. Waslander, "Coordinated landing of a quadrotor on a skid-steered ground vehicle in the presence of time delays," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, Sept 2011, pp. 4961–4966.

[17] M. Faessler, E. Mueggler, K. Schwabe, and D. Scaramuzza, "A monocular pose estimation system based on infrared leds," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, May 2014, pp. 907–913.

[18] S. Viollet, "Vibrating makes for better seeing: from the fly's micro eye movements to hyperacute visual sensors," *Frontiers in Bioengineering and Biotechnology*, vol. 2, no. 9, 2014.

[19] A. Manecy, S. Viollet, and N. Marchand, "Bio-inspired hovering control for an aerial robot equipped with a decoupled eye and a rate gyro," in *Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference on*, 2012, pp. 1110–1117.

[20] A. Manecy, R. Juston, N. Marchand, and S. Viollet, "Decoupling the eye: A key toward a robust hovering for sighted aerial robots," in *Chu, Qiping and Mulder, Bob and Choukroun, Daniel and Kampen, Erik-Jan and Visser, Coen and Looye, Gertjan*, N. Advances in Aerospace Guidance and Control, Eds. Springer Berlin Heidelberg, 2013, pp. 317–336.

[21] J. Gomez-Balderas, G. Flores, L. García Carrillo, and R. Lozano, "Tracking a ground moving target with a quadrotor using switching control," *Journal of Intelligent & Robotic Systems*, vol. 70, no. 1-4, pp. 65–78, 2013.

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

[23] Sensefly, "https://www.sensefly.com/drones/exom.html," 2014. [Online]. Available: https://www.sensefly.com/drones/exom.html

[24] L. Kerhuel, S. Viollet, and N. Franceschini, "Steering by gazing: An efficient biomimetic control strategy for visually guided micro aerial vehicles," *Robotics, IEEE Transactions on*, vol. 26, no. 2, pp. 307–319, April 2010.

[25] A. Manecy, N. Marchand, and S. Viollet, "Hovering by Gazing: a Novel Strategy for Implementing Saccadic Flight-Based Navigation in GPS-Denied Environments," *International Journal of Advanced Robotic Systems*, vol. 11, no. 66, p. , april 2014.

[26] L. Kerhuel, S. Viollet, and N. Franceschini, "The VODKA sensor: A bio-inspired hyperacute optical position sensing device," *Sensors Journal, IEEE*, vol. 12, no. 2, pp. 315–324, Feb 2012.

[27] L. Kerhuel, "Capteurs optiques minimalistes et réflexes oculomoteurs biomimétiques. applications à la robotique aérienne," Ph.D. dissertation, Information, Systèmes, Structures, Université Montpellier II, Dec 2009.

[28] R. Juston, L. Kerhuel, N. Franceschini, and S. Viollet, "Hyperacute edge and bar detection in a bioinspired optical position sensing device," *Mechatronics, IEEE/ASME Transactions on*, vol. 19, no. 3, pp. 1025–1034, June 2014.

[29] R. Juston, "De l'œil élémentaire à l'œil composé artificiel : application à la stabilisation visuelle en vol stationnaire," Ph.D. dissertation, Université d'Aix-Marseille, Nov 2013.

[30] F. Colonnier, A. Manecy, R. Juston, H. Mallot, R. Leitel, D. Floreano, and S. Viollet, "A small-scale hyperacute compound eye featuring active eye tremor: application to visual stabilization, target tracking, and short-range odometry," *Bioinspiration & Biomimetics*, vol. 10, no. 2, p. 026002, 2015.

[31] S. Viollet and N. Franceschini, "A high speed gaze control system based on the vestibulo-ocular reflex," *Robotics and Autonomous systems*, vol. 50, pp. 147–161, 2005.

[32] A. Manecy, R. Juston, N. Marchand, and S. Viollet, "Decoupling the eye: a key toward a robust hovering for sighted aerial robots," in *EuroGNC 2013, 2nd CEAS Specialist Conference on Guidance, Navigation & Control*, Delft, Netherland, April 2013.

[33] A. Manecy, N. Marchand, F. Ruffier, and S. Viollet, "X4-mag: a low-cost open-source micro-quadrotor and its linux-based controller," *International Journal of Micro Aerial Vehicles*, June 2015.

[34] A. Manecy, "RT-MaG Project, http://www.gipsa-lab.fr/projet/RT-MaG/," Jan. 2014. [Online]. Available: http://www.gipsa-lab.fr/projet/RT-MaG/

[35] M. Ruffo, M. D. Castro, L. Molinari, R. Losito, A. Masi, J. Kovermann, and L. Rodrigues, "New infrared time-of-flight measurement sensor for robotic platforms," *IMEKO TC4 Int. Symposium and Int. Workshop on ADC Modelling and Testing*, pp. 13–18, 2014.

[36] M. Faessler, F. Fontana, C. Forster, and D. Scaramuzza, "Automatic re-initialization and failure recovery for aggressive flight with a monocular vision-based quadrotor," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2015.

[37] D. Floreano, R. Pericet-Camara, S. Viollet, F. Ruffier, A. Brückner, R. Leitel, W. Buss, M. Menouni, F. Expert, R. Juston, M. K. Dobrzynski, G. L'Eplattenier, F. Recktenwald, H. A. Mallot, and N. Franceschini, "Miniature curved artificial compound eyes," *Proceedings of the National Academy of Sciences*, vol. 110, no. 23, pp. 9267–9272, 2013.