Visual homography-based pose estimation of a quadrotor using spectral features

Gastón Araguás, Claudio Paz, Gonzalo Perez Paina, Luis Canali
Centro de Investigación en Informática para la Ingeniería (CIII),
Universidad Tecnológica Nacional, Facultad Regional Córdoba,
Maestro Lopez S/N, Córdoba, Argentina.
Email: {garaguas,cpaz,gperez,lcanali}@frc.utn.edu.ar

***Abstract*—Pose estimation of Unmanned Aerial Vehicles (UAV) using cameras is currently a very active task in computer and robotic vision. This is mainly because of the use of robots in GPS-denied environments. However, the use of visual information for ego-motion estimation presents several difficulties, such as features search, data association, inhomogeneous features distribution in the image. This work addresses these issues by the use of the so-called spectral features, and a down-looking monocular camera rigidly attached to a quadrotor. We propose a visual position and orientation estimation algorithm based on the discrete homography constraint induced by the presence of planar scenes. This homography constraint results more appropriate than the well-known epipolar constraint, which vanishes for a zero translation and loses rank in the case of planar scenes. The pose estimation algorithm is tested in a simulated dataset and compared with the corresponding ground truth.**

Keywords*—*Motion estimation, Quadrotor, spectral features, discrete homography.

I. Introduction

In the last years quadrotors have gained popularity in entertainment, aero-shooting and many other civilian or military applications, mainly because of their low cost, great controllability, etc. Between other tasks, they are a good choice for operation at low altitude, in cluttered scenarios or even for indoor applications. Such environments limit the use of GPS or compass measurements which are indeed excellent options for attitude determination in wide open outdoor areas [11], [1]. These constraints have motivated, over the last years, the extensive use of on-board cameras as a main sensor for state estimation [13], [14], [4]. In this context, we present a new approach to estimate the self-motion of a quadrotor in indoor environments for stationary flights, using a down-looking camera for translation and rotation calculation. Stationary flights or hovering is a particular flight mode which consists in achieving that the six-degrees-of-freedom of the robot remain fixed around a state as stable as possible. This is an advantageous condition to point a camera downwards and to compute correspondences between images using homography. As a continuation of the work presented in [2], we propose the utilization of a fixed number of patches distributed on each image of the sequence to determine the self-motion of the camera, based on the plane-induced homography that relates the patches in two consecutive images. The pose of the camera (and UAV) is estimated in a "dead-reckoning" way, performing a time integration of self-motion parameters determined between images. We concentrate in the XY-position and the heading angle estimation in order to fuse these parameters with the on-board IMU and altimeter sensors measurements. The camera self-motion is estimated using the homography induced by the (assumed flat) floor, and the corresponding points are obtained on the frequency domain. This spectral information corresponds to an image patch which we call spectral feature [2], [3]. These kind of features perform better than the interest points based on the image intensity when observing a floor with homogeneous texture. Moreover, because their position in the image plane is previously selected, they are always well distributed.

The paper is organized as follows: Section II gives an overview of the related works about the image-to-image transformation estimation. Section III details the homography-based pose estimation, with a review of the so-called plane-induced homography. In this section the homography decomposition used to obtain the translation and rotation of the camera is also presented; and in order to estimate the homography the so-called spectral features are introduced in Subsection III-C. The implementation details and the algorithms are presented in Section IV. Results of the estimation parameters are presented in Section V, and finally Section VI remarks the conclusions and future work.

II. Related work

A number of spatial and frequency domain approaches have been proposed to estimate the image-to-image transformation between two views of a planar scene. Most of them are limited to similarity transformations. Spatial domain methods need corresponding points, lines, conics, etc. [6], [8], [9], whose identification in many practical situations is non-trivial, thereby limiting their applicability. Scale, rotation, and translation invariant features have been popular facilitating recognition under these transformations. Geometry of multiple views of the same scene has been a subject of extensive research over the past decade. Important results relating corresponding entities such as points and lines can be found in [6] [8]. Recent work has also focused on more complex entities such as conics and higher-order algebraic curves [9]. However, these approaches depend on extracting corresponding entities such as points, lines or contours and do not use the abundant information present in the form of the intensity values in the multiple views of the scene. Frequency domain methods are in general superior to methods based on spatial features [8] because the entire image information is used for matching. They also avoid the crucial issues regarding the selection of the best features.

Our work proposes the use of a fixed number of patches distributed on each image of the sequence to determine the pose change of a moving camera. The visual self-motion measurement is estimated using the homography induced by the (assumed flat) floor, and the corresponding points are obtained on the frequency domain. This spectral information corresponds to an image patch which we call spectral feature. These kind of features perform better than the interest points based on the image intensity when observing a floor with homogeneous texture. Moreover, because their position in the image plane is previously selected, they are always well distributed.

The transformation between two images taken from different views (with a moving camera) contains information about the spatial transformation of the views, or the camera movement. Considering a downward-looking camera, and assuming that the floor is a planar surface, all the space points imaged by the camera are coplanar and there is a homography between the world and the image planes. Under this constraint, if the camera center moves, the images taken from different points of view are also related by a homography. The spatial transformation that relates both views can be completely determined from this homography between images.

III. Homography-based pose estimation

The visual pose estimation is based on the principle that two consecutive images of a planar scene are related by a homography. The planar scene corresponds to the floor surface, which is assumed to be relatively flat, observed by the down-looking camera on the UAV. The spatial transformation of the camera, and therefore of the UAV, is encoded in this homography. Knowing the homography matrix that relates both images, the transformation parameters that describe the camera rotation and translation can be obtained.

In order to determine the homography induced by the planar surface, a set of corresponding points on two consecutive images must be obtained. This process is performed selecting a set of features in the first image and finding the corresponding set of features in the second one. Then, the image coordinates of each feature in both images conform the set of corresponding image points needed to calculate the homography.

The image features used in our approach are the so-called spectral features, a Fourier domain representation of an image patch. Selecting a set of spectral features in both images (the same number with the same size, and position), the displacement of each feature can be obtained using the Fourier shift theorem. This displacement, in addition to the feature center, determines the correspondence between features in both images.

A. Review of plane-induced homography

Given a 3D scene point $\mathbf{P}$, and two coordinate systems, CS_A and CS_B, the coordinates of the point $\mathbf{P}$ on each one can be denoted by $\mathbf{X}_A$ and $\mathbf{X}_B$ respectively. If $\boldsymbol{R}_A^B \in SO(3)$ is the rotation matrix that changes the representation of a point in CS_A to CS_B, and $\mathbf{T}_B \in \mathbb{R}^{3\times1}$ is the translation vector of the origin of CS_A w.r.t CS_B (expressed in CS_B), then the representations of the point $\mathbf{P}$ relate each other as

$$\mathbf{X}_B = \boldsymbol{R}_A^B \mathbf{X}_A + \mathbf{T}_B. \tag{1}$$

We suppose now that the point $\mathbf{P}$ belongs to a plane π, denoted in the coordinate system CS_A by its normal $\mathbf{n}_A$. Therefore, the following plane equation holds

$$\frac{(\mathbf{n}_A)^T \mathbf{X}_A}{d_A} = 1, \tag{2}$$

where d_A is the distance of the plane π to CS_A. Plugging (2) into (1) we have

$$\mathbf{X}_B = \left(\boldsymbol{R}_A^B + \frac{\mathbf{T}_B}{d_A} (\mathbf{n}_A)^T \right) \mathbf{X}_A = \boldsymbol{H}_A^B \mathbf{X}_A, \tag{3}$$

with

$$\boldsymbol{H}_A^B \doteq \left(\boldsymbol{R}_A^B + \frac{\mathbf{T}_B}{d_A} (\mathbf{n}_A)^T \right). \tag{4}$$

The matrix $\boldsymbol{H}_A^B$ is a *plane-induced homography*, in this case induced by the plane π. As can be seen, this matrix encodes the transformation parameters that relates both coordinates systems, ($\boldsymbol{R}_A^B$ and $\mathbf{T}_B$), and the structure parameters of the environment ($\mathbf{n}_A$ and d_A).

Considering now a moving camera associated to the coordinate system CS_A at time t_A and by CS_B at time t_B, according to the pin-hole camera model the relation between the 3D points and their projections are given by

$$\lambda_A \mathbf{x}_A = \mathbf{X}_A; \qquad \lambda_B \mathbf{x}_B = \mathbf{X}_B \tag{5}$$

where $\lambda_A \in \mathbb{R}^+$ and $\lambda_B \in \mathbb{R}^+$. Using (5) in equation (3) we have

$$\lambda_B \mathbf{x}_B = H_A^B \lambda_A \mathbf{x}_A \Rightarrow \qquad \mathbf{x}_B = \lambda H_A^B \mathbf{x}_A. \tag{6}$$

Given that both vectors $\mathbf{x}_B$ and $\lambda H_A^B \mathbf{x}_A$ have the same direction

$$\mathbf{x}_B \times \lambda H_A^B \mathbf{x}_A = \hat{\boldsymbol{x}}_B H_A^B \mathbf{x}_A = 0, \tag{7}$$

with $\hat{\boldsymbol{x}}_B$ the skew-symmetric matrix associated to $\mathbf{x}_B$. The equation (7) is known as the *planar epipolar restriction*, and holds for all 3D points belonging to the plane π. Assuming that the camera is pointing to the ground (downward-looking camera) and that the scene structure is approximately a planar surface, all the 3D points “seen” by the camera will hold this restriction.

The homography $\boldsymbol{H}_A^B$ represents the camera coordinate systems transformation between instant t_A and t_B: hence, it contains the information of the camera rotation and translation. It can be estimated knowing more than four corresponding points between two images. These correspondences are calculated in the spectral domain, by means of the *spectral features*. The process is detailed in subsection III-C.

B. Homography decomposition

Following [12] we can decompose $\boldsymbol{H}$ in order to obtain a non-unique solution (exactly four different solutions) $\left\{\boldsymbol{R}_i, \mathbf{n}_i, \frac{\mathbf{T}_i}{d_i}\right\}$, and then adding some extra data for disambiguation we arrive to the appropriate $\left\{\boldsymbol{R}_A^B, \mathbf{n}_A, \frac{\mathbf{T}_B}{d_A}\right\}$ solution.

1) Normalization: Given that the planar epipolar constraint ensures equality only in the direction of both vectors (equation (7)), what we really have after the homography estimation is $\lambda\boldsymbol{H}$, that is[1]

$$\boldsymbol{H}_\lambda = \lambda\boldsymbol{H} = \lambda\left(\boldsymbol{R} + \frac{\mathbf{T}}{d}\mathbf{n}^T\right). \tag{8}$$

The unknown factor λ included in $\boldsymbol{H}_\lambda$ can be found as follows. Consider the product

$$\boldsymbol{H}_\lambda^T\boldsymbol{H}_\lambda = \lambda^2\left(\boldsymbol{I} + \boldsymbol{Q}\right) \tag{9}$$

with $\boldsymbol{I}$ the identity, $\boldsymbol{Q} = \mathbf{a}\mathbf{n}^T + \mathbf{n}\mathbf{a}^T + ||\mathbf{a}||^2\mathbf{n}\mathbf{n}^T$ and $\mathbf{a} = \frac{1}{d}\boldsymbol{R}^T\mathbf{T} \in \mathbb{R}^{3\times1}$. The vector $\mathbf{a}\times\mathbf{n}$, perpendicular to $\mathbf{a}$ and $\mathbf{n}$, is an eigenvector of $\boldsymbol{H}_\lambda^T\boldsymbol{H}_\lambda$ associated to the eigenvalue λ^2, being that

$$\boldsymbol{H}_\lambda^T\boldsymbol{H}_\lambda(\mathbf{a}\times\mathbf{n}) = \lambda^2(\mathbf{a}\times\mathbf{n}). \tag{10}$$

So, if λ^2 is an eigenvalue of $\boldsymbol{H}_\lambda^T\boldsymbol{H}_\lambda$, then $|\lambda|$ is a singular value of $\boldsymbol{H}_\lambda$. It is easy to show that $\boldsymbol{Q}$ in (9) has one positive, one zero and one negative eigenvalue, what means that λ^2 is the second ordered eigenvalue of $\boldsymbol{H}_\lambda^T\boldsymbol{H}_\lambda$ and $|\lambda|$ will be the second ordered singular value of $\boldsymbol{H}_\lambda$. That is, if $\sigma_1 > \sigma_2 > \sigma_3$ are the singular values of $\boldsymbol{H}_\lambda$, then

$$\boldsymbol{H} = \pm\frac{\boldsymbol{H}_\lambda}{\sigma_2} \tag{11}$$

To get the right sign of $\boldsymbol{H}$, the positive depth condition in (6) must be applied. In order to ensure that all the considered points are in front of the camera, all 3D points in plane π projected in the image plane must fulfill

$$(\mathbf{x}_B^j)^T\boldsymbol{H}\mathbf{x}_A^j = \frac{1}{\lambda_j} > 0, \quad \forall j = 1, 2, \ldots, n. \tag{12}$$

where $\left(\mathbf{x}_A^j, \mathbf{x}_B^j\right)$ are the projections of all points $\{\mathbf{P}\}_{j=1}^n$ lying on the plane π, at time t_A and t_B respectively.

2) Estimation of $\mathbf{n}$*:* The homography $\boldsymbol{H}$ induced by the plane π preserves the norm of any vector in the plane, i.e. given a vector $\mathbf{r}$ such that $\mathbf{n}^T\mathbf{r} = 0$, then

$$\boldsymbol{H}\mathbf{r} = \boldsymbol{R}\mathbf{r} \tag{13}$$

and therefore $||\boldsymbol{H}\mathbf{r}|| = ||\mathbf{r}||$. Consequently, knowing the space spanned by the vectors that preserve the norm under $\boldsymbol{H}$, the perpendicular vector $\mathbf{n}$ is also known.

The matrix $\boldsymbol{H}^T\boldsymbol{H}$ is symmetric, and therefore admits eigenvalue decomposition. Being $\sigma_1^2, \sigma_2^2, \sigma_3^2$ the eigenvalues and $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ the eigenvectors of $\boldsymbol{H}^T\boldsymbol{H}$, then

$$\begin{gathered}\boldsymbol{H}^T\boldsymbol{H}\mathbf{v}_1 = \sigma_1^2\mathbf{v}_1, \quad \boldsymbol{H}^T\boldsymbol{H}\mathbf{v}_2 = \mathbf{v}_2, \\ \boldsymbol{H}^T\boldsymbol{H}\mathbf{v}_3 = \sigma_3^2\mathbf{v}_3\end{gathered} \tag{14}$$

since by the normalization $\sigma_2^2 = 1$. That is, $\mathbf{v}_2$ is perpendicular to $\mathbf{n}$ and $\mathbf{T}$, so its norm is preserved under $\boldsymbol{H}$. From (14) it can be shown that the norm of the following vectors

$$\begin{aligned}\mathbf{u}_1 &\doteq \frac{\sqrt{1-\sigma_3^2}\mathbf{v}_1 + \sqrt{\sigma_1^2-1}\mathbf{v}_3}{\sqrt{\sigma_1^2-\sigma_3^2}}, \\ \mathbf{u}_2 &\doteq \frac{\sqrt{1-\sigma_3^2}\mathbf{v}_1 - \sqrt{\sigma_1^2-1}\mathbf{v}_3}{\sqrt{\sigma_1^2-\sigma_3^2}}\end{aligned} \tag{15}$$

is preserved under $\boldsymbol{H}$ too, as well as all vectors in the subspaces spanned by

$$S_1 = \text{span}\left\{\mathbf{v}_2, \mathbf{u}_1\right\}, \quad \text{S}_2 = \text{span}\left\{\mathbf{v}_2, \mathbf{u}_2\right\} \tag{16}$$

Therefore, there exist two possible planes that can induce the homography $\boldsymbol{H}$, π_1 and π_2, defined by the normal vectors to S_1 and S_2

$$\mathbf{n}_1 = \mathbf{v}_2 \times \mathbf{u}_1, \quad \mathbf{n}_2 = \mathbf{v}_2 \times \mathbf{u}_2. \tag{17}$$

3) Estimation of $\boldsymbol{R}$*:* The action of $\boldsymbol{H}$ over $\mathbf{v}_2$ and $\mathbf{u}_1$ is equivalent to a pure rotation

$$\boldsymbol{H}\mathbf{v}_2 = \boldsymbol{R}_1\mathbf{v}_2, \quad \boldsymbol{H}\mathbf{u}_1 = \boldsymbol{R}_1\mathbf{u}_1 \tag{18}$$

since both vectors are orthogonal to $\mathbf{n}_1$. The rotation of $\mathbf{n}_1$ can be computed as

$$\boldsymbol{R}_1\mathbf{n}_1 = \boldsymbol{H}\mathbf{v}_2 \times \boldsymbol{H}\mathbf{u}_1. \tag{19}$$

From (18) and (19) we have

$$\boldsymbol{R}_1 = [\boldsymbol{H}\mathbf{v}_2, \boldsymbol{H}\mathbf{u}_1, \boldsymbol{H}\mathbf{v}_2 \times \boldsymbol{H}\mathbf{u}_1][\mathbf{v}_2, \mathbf{u}_1, \mathbf{n}_1]^T. \tag{20}$$

Considering now the set $\{\mathbf{v}_2, \mathbf{u}_2, \mathbf{n}_2\}$, in the same way we arrive to

$$\boldsymbol{R}_2 = [\boldsymbol{H}\mathbf{v}_2, \boldsymbol{H}\mathbf{u}_2, \boldsymbol{H}\mathbf{v}_2 \times \boldsymbol{H}\mathbf{u}_2][\mathbf{v}_2, \mathbf{u}_2, \mathbf{n}_2]^T. \tag{21}$$

Once $\boldsymbol{R}$ and $\mathbf{n}$ are known, the estimation of $\frac{\mathbf{T}}{d}$ is direct. However, it must be noted that the term $\frac{\mathbf{T}}{d}\mathbf{n}^T$ in $\boldsymbol{H}$ introduces a sign ambiguity, since $\frac{\mathbf{T}}{d}\mathbf{n}^T = \frac{-\mathbf{T}}{d}(-\mathbf{n}^T)$, therefore the number of possible solutions rises to four,

$$\begin{gathered}\left\{\boldsymbol{R}_1, \mathbf{n}_1, \frac{\mathbf{T}_1}{d_1}\right\}, \quad \left\{\boldsymbol{R}_1, -\mathbf{n}_1, \frac{-\mathbf{T}_1}{d_1}\right\}, \\ \left\{\boldsymbol{R}_2, \mathbf{n}_2, \frac{\mathbf{T}_2}{d_2}\right\}, \quad \left\{\boldsymbol{R}_2, -\mathbf{n}_2, \frac{-\mathbf{T}_2}{d_2}\right\}\end{gathered} \tag{22}$$

In order to ensure that the plane inducing the homography $\boldsymbol{H}$ appears in front of the camera, each normal vector $\mathbf{n}_i$ must fulfill $n_z > 0$, and therefore only two solutions remain. These two solutions are both possible physically, but given that most of the time the camera on the UAV is facing-down, we choose the solution with the n_z component closest to zero.

[1] To avoid the abuse of notation we do not use here the sub and supra indexes A and B that refers to the corresponding coordinate systems.

C. Spectral features correspondence

The so-called spectral feature refers to the Fourier domain representation of an image patch of $2^n \times 2^n$, where $n \in \mathbb{N}^+$ is set accordingly to the allowed image displacement [2]. The power of 2 of this patch size is selected based on the efficiency of the Fast Fourier Transform (FFT) algorithm. The number and position of spectral features in the image are set beforehand. Even though a minimum of four points are needed to estimate the homography, a higher number of features are used to increase the accuracy and the RANSAC algorithm [7] is used for outliers elimination.

Consider two consecutive frames, where spectral features on each image were computed. To determine the correspondence between features is equivalent to determine the displacement between them. This displacement can be obtained using the spectral information by means of the Phase Correlation Method (PCM) [10]. This method is based on the Fourier shift theorem, which states that the Fourier transforms of two identical but displaced images differ only in a phase shift.

Given two images i_A and i_B of size $N \times M$ differing only in a displacement (u, v), their Fourier transforms are related by

$$I_A(\omega_x, \omega_y) = e^{-j(u\omega_x + v\omega_y)} I_B(\omega_x, \omega_y), \tag{23}$$

and therefore this displacement can be obtained from (23) using the cross-power spectrum of the given transformations I_A and I_B

$$\frac{I_A I_B^*}{|I_A||I_B^*|} = e^{-j(u\omega_x + v\omega_y)}. \tag{24}$$

The inverse transform of (24) is an impulse located exactly in (u, v), which represents the displacement between the two images

$$\mathcal{F}^{-1}[e^{-j(u\omega_x + v\omega_y)}] = \delta(x - u, y - v). \tag{25}$$

Using the discrete FFT (Fast Fourier Transform) algorithm instead of the continuous version, the result will be a pulse signal centered in (u, v) [15].

D. Homography estimation

The previous subsection describes how to calculate the displacement between two images using PCM. Applying this method to each image patch pair, the displacement between spectral features is determined. The set of corresponding points required to estimate the homography can be constructed with the patch centers of the first image and the displaced patch centers of the second one, that is

$$\{\mathbf{x}_{A_i} \leftrightarrow \mathbf{x}_{A_i} + \Delta\mathbf{d}_i = \mathbf{x}_{B_i}\} \tag{26}$$

where $\Delta\mathbf{d}_i$ represents the displacement between the i-th spectral feature, and $\mathbf{x}_{A_i}$ the center of the i-th spectral feature in the CS_A. This is schematically shown in the zoomed area of Fig. 1. As shown in section III-A, this set of corresponding points is related by a homography from which, using linear methods plus nonlinear optimization, the associated homography matrix can be computed [8].

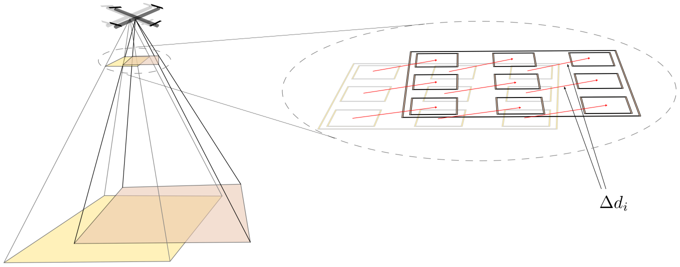

Fig. 1. Estimation of the rotation and translation between two consecutive images based on spectral features.

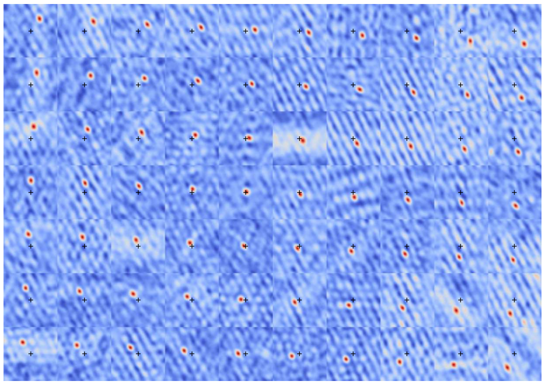

Fig. 2. Displacements between patches.

In Fig. 2 a real set of spectral features is shown, where the black crosses represent each patch center and the yellow circles represent the output of PCM.

It is important to note that the number, size, and position of spectral features are set beforehand: therefore, neither a search nor a correspondence process needs to be performed. Moreover, due to the fact that the spectral features use the frequency spectrum of the image intensity as feature descriptor, they result to be more robust than the gradient-based features, which in general work only in presence of corners in the image.

IV. Implementation

Summarizing, Alg. 1 shows the proposed procedure to estimate the position and orientation, Alg. 2 shows the procedure to determine the displacement between patches, and in Alg. 3 the homography decomposition process is detailed.

Algorithm 1 Position and orientation estimation.

```
function POSEESTIMATION($i_t, i_{t-1}$)
    Extract patches $p_{i\,t}$ and $p_{i\,t-1}$ from $i_t$ y $i_{t-1}$
    for all $\{p_{i\,t}, p_{i\,t-1}\}$ do
        $\Delta\mathbf{d}_i \leftarrow$ FINDDISPLACEMENT($p_{i\,t}, p_{i\,t-1}$)
        $\mathbf{x}_{i\,t} \leftarrow \mathbf{x}_{i\,t-1} + \Delta\mathbf{d}_i$
    end for
    $H_\lambda \leftarrow$ FINDHOMOGRAPHY($\mathbf{x}_{i\,t}, \mathbf{x}_{i\,t-1}$)
    $R$, $\mathbf{n}$, $\mathbf{T/d} \leftarrow$ GETRTN($H_\lambda$)
    return $R$, $\mathbf{n}$, $\mathbf{T/d}$
end function
```

Algorithm 2 Patches displacement determination.

```
function FINDDISPLACEMENT($p_{i\,t}, p_{i\,t-1}$)
    $P_{i\,t} \leftarrow$ FASTFOURIERTRANSFORM($p_{i\,t}$)
    $P_{i\,t-1} \leftarrow$ FASTFOURIERTRANSFORM($p_{i\,t-1}$)
    $C \leftarrow$ CROSSPOWERSPECTRUM($P_{i\,t}, P_{i\,t-1}$)
    $r \leftarrow$ INVERSEFASTFOURIERTRANSFORM($c$)
    $\Delta\mathbf{d}_i \leftarrow \operatorname{argmax} r$
    return $\Delta\mathbf{d}_i$
end function
```

V. Results

The evaluation of the proposed visual pose estimation approach is performed with synthetic images obtained from a simulated quadrotor. In order to be able to generate a six degrees of freedom motion similar to a real quadrotor a simulated dynamic model was used. This allows to get the truth robot position and orientation which are then used to crop a sequence of images from a big one representing the observed flat surface. The ground truth pose is also used for evaluation purposes. The simulation of the quadrotor is based on Simulink, and the dynamic model is presented in [5]. Figure 3 shows an example of the path followed by the quadrotor, used to generate the synthetic dataset.

The path consists on a change of altitude followed by two loops maintaining constant radius. During the loops, the heading angle, also called yaw angle, was set to grow up to 2π radians.

The images were obtained from a *virtual* downward-looking camera following the path described above, cutting portions of 640×480 from a bigger image of uniform distributed noise in order to simulate a carpet. The virtual camera was configured with a pixel size of 5.6μm and a focal length of approx. 1mm. The algorithm was set with 42 patches of 128×128 pixels, equally distributed in the image.

In Figure 4 the estimated parameters together with the ground truth are shown. The graphic on the top shows the X position estimation of the UAV, which performs a total of $2.5m$ of change in the complete trajectory. The Y position estimation is plotted in the middle, and it has a similar behavior to the X one. As can be seen, the estimation error remains bounded in both axes all the time. The last graphic shows the yaw angle estimation, which follows the ground truth with a very small error.

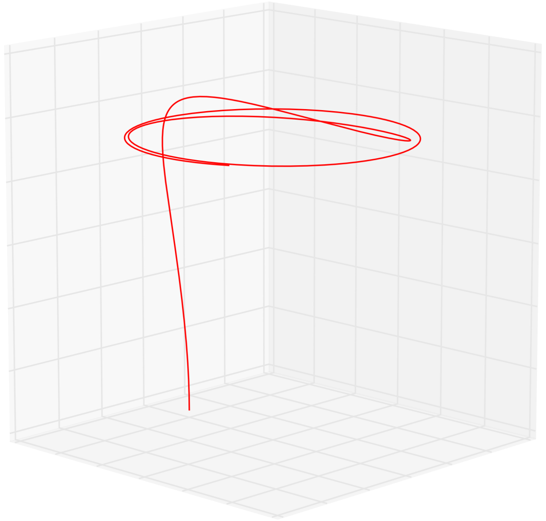

Fig. 3. Simulated position of a quadrotor with a six-degrees-of-freedom motion.

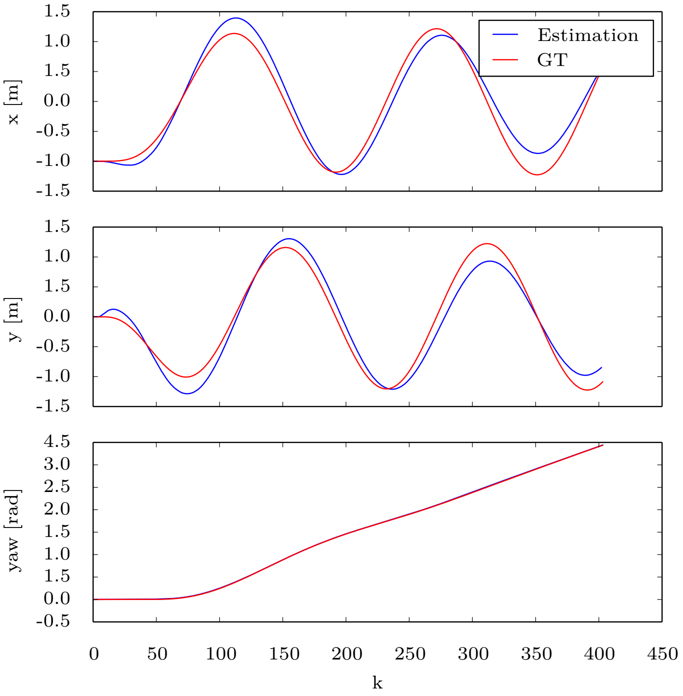

Fig. 4. Estimation of the XY-position and yaw angle of the UAV during a $20s$ flight.

VI. Conclusions

In this work a new approach for visual estimation of the pose change of a quadrotor with a down-looking camera was presented. The proposed algorithm is based on the plane-induced homography that relates two views of the floor. The downward-looking camera is used to estimate the corresponding points for the homography estimation based on spectral features. The main advantage of using spectral features as in this implementation, is that the typical correspondence process does not need to be performed.

The evaluation of the visual algorithm using a synthetic dataset has shown that the XY-position is estimated without significant absolute error, despite the typical accumulated error of the integration process. It is important to note that the view changes introduced by the orientation change (roll and pitch) over the flight did not induce any considerable error in the XY-position estimation. Likewise, the estimation of the heading (yaw) angle has shown to be accurate enough to be used in a IMU-camera fusion schema.

Algorithm 3 Homography matrix decomposition.

```
function GETRTN($H_\lambda$)
    $U_\lambda,\ \Sigma_\lambda,\ V_\lambda^T \leftarrow \text{SVDecomp}(H_\lambda)$
    $H \leftarrow H\lambda/\sigma_2$
    $U,\ \Sigma,\ V^T \leftarrow \text{SVDecomp}(H)$
    $\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3\end{bmatrix} \leftarrow V$
    $\mathbf{u}_1 \leftarrow \frac{\mathbf{v}_1\sqrt{1-\sigma_3^2}+\mathbf{v}_3\sqrt{\sigma_1^2-1}}{\sqrt{\sigma_1^2-\sigma_3^2}}$ ;    $\mathbf{u}_2 \leftarrow \frac{\mathbf{v}_1\sqrt{1-\sigma_3^2}-\mathbf{v}_3\sqrt{\sigma_1^2-1}}{\sqrt{\sigma_1^2-\sigma_3^2}}$
    $\mathbf{n}_1 \leftarrow \mathbf{v}_2 \times \mathbf{u}_1$ ;    $\mathbf{n}_2 \leftarrow \mathbf{v}_2 \times \mathbf{u}_2$
    Choose only the two physically possible solutions (this ensures that $\mathbf{n}_1$ and $\mathbf{n}_2$ have $n_z$ positive component)
    $U_1 \leftarrow \begin{bmatrix}\mathbf{v}_2 & \mathbf{u}_1 & \mathbf{n}_1\end{bmatrix}$ ;    $U_2 \leftarrow \begin{bmatrix}\mathbf{v}_2 & \mathbf{u}_2 & \mathbf{n}_2\end{bmatrix}$
    $W_1 \leftarrow \begin{bmatrix}H\mathbf{v}_2 & H\mathbf{u}_1 & H\mathbf{v}_2 \times H\mathbf{u}_1\end{bmatrix}$ ;    $W_2 \leftarrow \begin{bmatrix}H\mathbf{v}_2 & H\mathbf{u}_2 & H\mathbf{v}_2 \times H\mathbf{u}_2\end{bmatrix}$
    $R_1 \leftarrow W_1U_1^T$ ;    $R_2 \leftarrow W_2U_2^T$
    $T_1/d \leftarrow (H-R_1)\mathbf{n}_1$ ;    $T_2/d \leftarrow (H-R_2)\mathbf{n}_2$
    Choose the solution with $n_z$ of each normal plane vector closest to zero
    return $R$, $\mathbf{n}$, $\mathbf{T/d}$
end function
```

Acknowledgment

This work was partially funded by Universidad Tecnológica Nacional through the project "Fusión Sensorial para Estimación de Posición y Orientación 3D", UTN-PID-2155, and the Argentinian National Agency for the Advance in Science and Technology through the project "Autonomous Vehicle Guidance Fusing Low-cost GPS and other Sensors", PICT-PRH-2009-0136. Both currently under development at CIII, UTN, Córdoba, Argentina.

References

[1] M. Angermann, M. Frassl, M. Doniec, B.J. Julian, and P. Robertson. Characterization of the indoor magnetic field for applications in localization and mapping. In *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*, pages 1–9, 2012.

[2] Gastón Araguás, Claudio Paz, David Gaydou, and Gonzalo Perez Paina. Quaternion-based orientation estimation fusing a camera and inertial sensors for a hovering UAV. *Journal of Intelligent & Robotic Systems*, 77(1):37–53, August 2014.

[3] Gastón Araguás, Jorge Sánchez, and Luis Canali. Monocular visual odometry using features in the fourier domain. In *VI Jornadas Argentinas de Robótica*, Instituto Tecnológico de Buenos Aires, Buenos Aires, Argentina, 2010.

[4] Francisco Bonin-Font, Alberto Ortiz, and Gabriel Oliver. Visual navigation for mobile robots: A survey. *Journal of Intelligent & Robotic Systems*, 53(3):263–296, 2008.

[5] Peter Corke. *Robotics, Vision and Control*, volume 73 of *Springer Tracts in Advanced Robotics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[6] Olivier Faugeras and Quang-Tuan Luong. *The geometry of multiple images: the laws that govern the formation of multiple images of a scene and some of their applications*. MIT press, 2004.

[7] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.

[8] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

[9] Jeremy Yermiyahou Kaminski and Amnon Shashua. Multiple view geometry of general algebraic curves. *International Journal of Computer Vision*, 56(3):195–219, 2004.

[10] C. D. Kuglin and D. C. Hines. The phase correlation image alignment method. *Proc. Int. Conf. on Cybernetics and Society*, 4:163–165, 1975.

[11] Binghao Li, T. Gallagher, A.G. Dempster, and C. Rizos. How feasible is the use of magnetic field alone for indoor positioning? In *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*, pages 1–9, 2012.

[12] Yi Ma, Stefano Soatto, Jana Kosecká, and S. Shankar Sastry. *An Invitation to 3-D Vision: From Images to Geometric Models*. Springer, New York, NY, November 2010.

[13] D. Scaramuzza, M.C. Achtelik, L. Doitsidis, F. Friedrich, E. Kosmatopoulos, A. Martinelli, M.W. Achtelik, M. Chli, S. Chatzichristofis, L. Kneip, D. Gurdan, L. Heng, Gim Hee Lee, S. Lynen, M. Pollefeys, A. Renzaglia, R. Siegwart, J.C. Stumpf, P. Tanskanen, C. Troiani, S. Weiss, and L. Meier. Vision-controlled micro flying robots: From system design to autonomous navigation and mapping in GPS-Denied environments. *IEEE Robotics Automation Magazine*, 21(3):26–40, sep 2014.

[14] Stephan Weiss, Markus W. Achtelik, Simon Lynen, Michael C. Achtelik, Laurent Kneip, Margarita Chli, and Roland Siegwart. Monocular vision for long-term micro aerial vehicle state estimation: A compendium. *Journal of Field Robotics*, 30(5):803–831, September 2013.

[15] Barbara Zitová and Jan Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, October 2003.