Global Descriptors for Visual Pose Estimation of a Non-Cooperative Target in Space Rendezvous

Anthea Comellini, Jerome Le Ny, Emmanuel Zenou, Christine Espinosa, Vincent Dubanchet

Global Descriptors for Visual Pose Estimation of a Non-Cooperative Target in Space Rendezvous

Anthea Comellini, Jerome Le Ny, *Senior Member, IEEE*, Emmanuel Zenou, Christine Espinosa, and Vincent Dubanchet

Abstract

This paper proposes methods based on global descriptors to estimate the pose of a known object using a monocular camera, in the context of space rendezvous between an autonomous spacecraft and a non-cooperative target. These methods estimate the pose by *detection*, i.e., they require no prior information about the pose of the observed object, making them suitable for initial pose acquisition and the monitoring of faults in other on-board estimators. An approach is presented to fully retrieve the object's pose using a pre-computed set of invariants and geometric moments. Three classes of global invariant features are analyzed, based on complex moments, Zernike moments and Fourier descriptors. The robustness of the different invariants is tested under various conditions and their performance is discussed and compared. The method offers a fast and robust solution for pose estimation by detection, with a low computational complexity that is compatible with space-qualified processors.

I. Introduction

During an autonomous space rendezvous (RDV) with a non-cooperative target, i.e., a target that does not assist the chaser spacecraft (S/C) in acquisition, tracking and rendezvous operations [1],

PhD student, ISAE-SUPAERO and Institut Clément Ader ICA, 31400 Toulouse, France, and Thales Alenia Space, 06150 Cannes, France; anthea.comellini@isae-supaero.fr

Associate professor, Polytechnique Montreal, QC H3T 1J4, Canada; jerome.le-ny@polymtl.ca

Associate professor, ISAE-SUPAERO, 31400 Toulouse, France; emmanuel.zenou@isae-supaero.fr

Research professor, Institut Clément Ader ICA, Université de Toulouse, ISAE-SUPAERO, IMT MINES ALBI, UTIII, INSA, CNRS, 31400 Toulouse, France; christine.espinosa@isae-supaero.fr

PhD and GNC engineer at Thales Alenia Space, Cannes, France, 06150; vincent.dubanchet@thalesaleniaspace.com

the S/C must estimate the target's state on-board autonomously. Autonomous RDV operations require accurate, up-to-date measurements of the relative pose (i.e., position and attitude) of the target. Since passive camera sensors have a small form factor and a low power consumption, they can be easily integrated on a S/C, without affecting its design and its power budget. For this reason, the combination of camera sensors with tracking algorithms can provide a cost effective solution. In this article, we will assume to rely on a single camera (i.e., monocular vision) for the pose estimation. Indeed, monocular vision has advantages compared to stereo-camera configurations, because single camera systems are less complex and have a much larger operational range, which is not limited by the size of the platform.

Image-based pose estimation methods can be classified into *non-model-based* and *model-based* techniques [2]. Non-model-based techniques do not assume any *a priori* knowledge of the tracked object's shape, texture and other visual attributes. They simultaneously estimate the object's pose while reconstructing its visual model [3], [4], [5]. On the other hand, model-based techniques take advantage of *a priori* knowledge about the object whose pose is to be estimated. This knowledge can be in the form of visual features such as markers, or in the form of a 3D geometric description of the object. Since in the majority of rendezvous scenarios the target spacecraft is a known object, we focus here on model-based techniques. Model-based 3D tracking methods can be classified as *frame-by-frame tracking* or *tracking-by-detection* [6]. In frame-by-frame tracking, the object's pose retrieved from a frame is used as a prior for a local search of the pose in the following frame. This recursive approach makes image feature identification and matching relatively easy [6]. However, these methods require initialization and can diverge in the presence of local minima. For this reason, a recursive tracking algorithm must be complemented with a tracking-by-detection algorithm to enable initial pose acquisition and fault detection. In a tracking-by-detection algorithm, the pose is retrieved by exploiting a-priori information on the geometry and appearance of the tracked object, but with no knowledge of the pose at previous instants. Since space-qualified microprocessors have relatively low computational resources, the tracking-by-detection algorithm must be computationally inexpensive. Moreover, it has to cope with the peculiarities of the RDV problem, such as harsh illumination conditions or the presence of textureless and reflective materials on the target.

A. Monocular Tracking-by-Detection

Monocular pose estimation by detection can be performed using geometric methods or by template matching. In geometric methods, the observed 2D features in the input image are matched with a database of features computed offline. Then the pose is retrieved by solving the Perspective-n-Point (PnP) problem. Geometric approaches based on local feature are proposed for S/C pose acquisition in [7], [8], [9]. However, methods relying on feature detection lack robustness in the adverse illumination conditions encountered by spaceborne systems. In template matching approaches, a *training* set of views of the object is acquired offline to generate a database of templates that are compared at run-time with the input image. Classical template matching approaches compare the pixels intensities of the templates and the input image according to similarity measures or alignment functions [10]. However, these methods are computationally expensive and lack robustness to illumination changes [10]. Other approaches rely on templates of local features, such as image gradient orientations (e.g., [11], [12]) or binary templates of the extracted edges. An edge-based template matching approach relying on a similarity measure derived from *Chamfer Matching* [13] and on an unsupervised clustering technique is proposed for S/C pose estimation in [10]. The template matching approaches described are computationally complex because of the evaluation of a large number of possible pose hypotheses and real-time may not be achievable on space qualified processors. They might be appropriate for pose initialization, but cannot be used as a backup algorithm to help detecting divergence in the recursive tracking algorithm. For this reason, recent work has focused on using Convolutional Neural Networks (CNNs) for pose estimation directly from greyscale images in a end-to-end fashion [14], [15]. However, the reported accuracy is currently still lower than that of geometric methods [7]. Hybrid approaches, where CNNs are used to extract keypoints (i.e., local features) from the image while the PnP solver is used to compute the pose, have recently shown very good accuracy with synthetic images [16]. However, this still raises the issue of relying on local feature detection.

The approach proposed in this paper is to rely on template matching with global features instead of local ones. Global features such as image moments or Fourier descriptors provide a low-dimensional representation of the target's silhouette on a binary image. As for a typical appearance-based template approach, this representation can be matched to its nearest neighbor in a database constructed offline from a sufficiently rich sample of possible poses. However, due

to the small amount of data stored for each view, the search for the optimal matching remains computationally inexpensive, resulting in a fast estimation procedure that can be run in parallel of a frame-to-frame tracking algorithm. If global features are computed from binary silhouette images, they become independent from illumination conditions. However, in this case, the pose estimation algorithm should be complemented with a segmentation algorithm, especially when complex background such as the Earth is present in the image.

B. Pose Estimation with Global Descriptors

Early attempts to use global features for shape recognition were motivated by aircraft identification applications and relied on Hu's moment invariants [17], [18], [19], [20] and Fourier descriptors [21], [18], [22], [19]. Dudani proposed in [17] to use these features not only for aircraft recognition but also for pose estimation, inspiring follow-up work in [18], [22], [19], [20]. Unfortunately, Hu's invariants, although very commonly used, are now known to form an incomplete set, see Section IV, which results in limited discrimination capabilities. During the past two decades, progress in pattern recognition led to the development of more powerful sets of rotation-invariant global features, which however were never tested for the pose estimation problem. In this paper, we consider complex moments (CMs), Zernike moments (ZMs) and Fourier descriptors (FDs). CMs directly improve on Hu's invariants [23], see Section IV-A. ZMs form a set of orthogonal moments, see Section IV-B, with advantages in terms of information redundancy and image representation capabilities [24], [25]. Some authors claim that Pseudo-ZM [26] are more robust than ZM with respect to additive noise, but their independence (see Section IV) is considered questionable [23]. ZMs were used in [27] to determine the orientation of a S/C from silhouette images, as we do here. However, the authors did not fully exploit the power of the ZMs since they do not use the rotation invariance to reduce the complexity of the problem, as we explain in Section III. In [28] the amplitude of ZMs up to the 29^{th} order is used as rotation invariant to estimate the pose of airplanes, but the use of only amplitudes leads to an incomplete set of descriptors, which results in limited discrimination capabilities, as we explain in Section IV-B. Moreover, in [28] the in-plane rotation (see Section III) is retrieved by comparing the phase of the ZMs in the database with the phase of the ZMs computed on the observed image. This approach leads to a doubling of the dimension of the descriptor database. In Section III, we describe a method to retrieve the in-plane rotation that requires storing only one descriptor for each database view. The amplitude of the Pseudo-ZMs is also used as rotation invariant in [29],

but this work proposes a method to compute only two of the six degrees-of-freedom (DOF) of the pose. Finally, Pseudo-ZMs are used in [30] for aircraft pose estimation from contour images. First, the authors determine with Pseudo-ZMs two of the three attitude angles of the observed object. Then, shape context descriptors are used to retrieve the remaining DOF. However, as we details in Section III, the remaining DOF can be computed using geometric moments, resulting in a simpler and faster estimation method.

Moment-based invariants are generally computed using the whole silhouette of the observed object, since the performance of boundary moments quickly deteriorates in the presence of noise and discretization effects on the images [18], [31]. Thus CMs and ZMs are generally sensitive to the distribution of the "mass" in the image silhouette. On the other hand, FDs are computed only from the contours of the observed object, hence are more sensitive to changes in the object boundaries [18]. This motivates our goal of comparing Fourier descriptors and moment-based invariants. The contributions of this paper can be summarized as follows. After presenting the problem statement in Section II, we develop in Section III a tracking-by-detection algorithm for the pose estimation of a S/C of known geometry, based on global features computed from a single binary image. First, the method determines two of the three Euler angles that describe the attitude of the S/C using rotation invariant features. Then, the remaining degrees of freedom, i.e., the relative translation and the in-plane rotation, are retrieved using geometric moments (GMs). In Section IV the theory underlying the computation of rotation invariants based on CMs, ZMs and FDs is recalled, and we propose a novel method to compute rotation invariant FDs. Then, Section V analyzes and compares the performance of these three sets of global descriptors, characterizing their behavior in different conditions typically encountered in operational RDV scenarios. To the best of our knowledge, no prior work provides such an evaluation of the performance of modern global descriptors for the problem of detecting and estimating the pose of a $3D$ object.

II. Problem statement

The problem considered in this paper is to estimate the pose of an object of known geometry from a single bi-dimensional image. No prior information about the object's pose is known. We assume that the input of the estimation procedure is a binarized image of the object's silhouette and we do not address in this paper the task of foreground-background segmentation. However, for operational scenarios such as a RDV with a space debris, the object can be approached

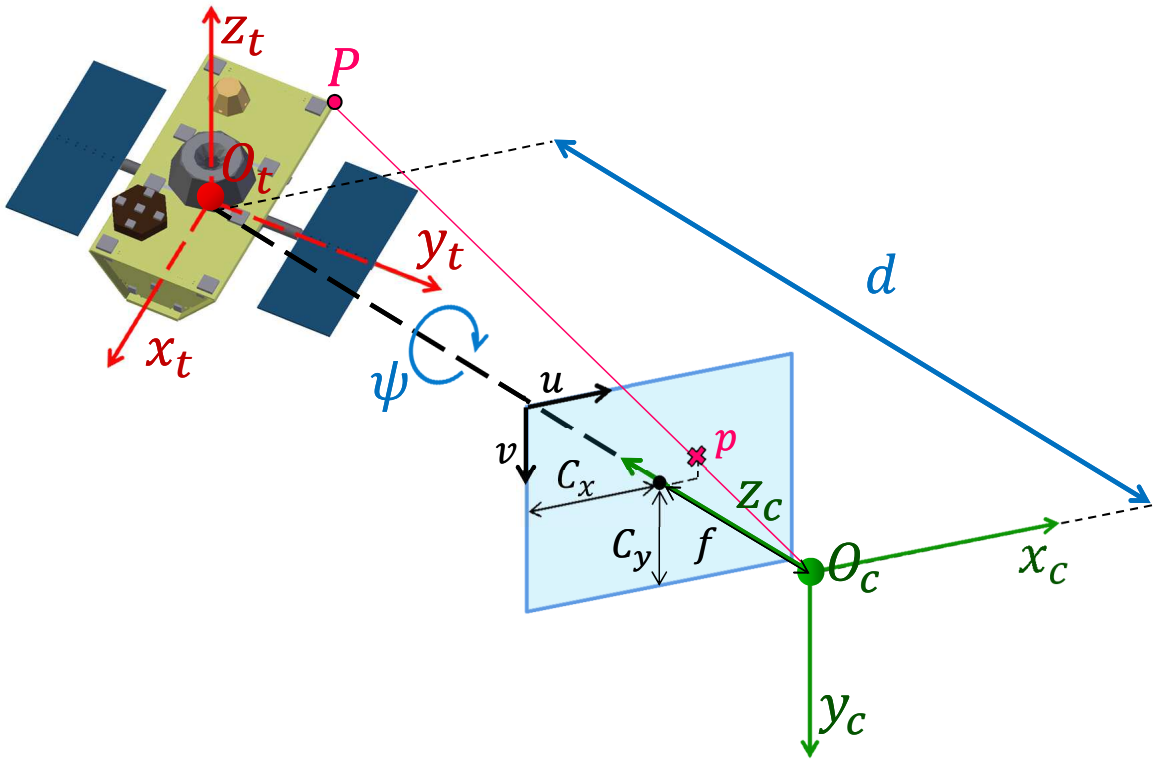


Fig. 1. Schematic representation of the pose estimation problem using a monocular image

from a lower orbit, keeping the Earth out of the field of view (FOV) of the camera. Then, the silhouette of the object can be easily retrieved from a multi-spectral image after a simple thresholding. Fig. 1 shows a schematic representation of the pose estimation problem. Let P^t be the coordinates of a point P expressed in the target reference frame (RF) t, which is centered at the target center of mass (COM). The coordinates of P, expressed in the camera RF, are:

$$P^c = [x^c,\, y^c,\, z^c]^T = R_{c-t}\, P^t + t^c_{O_c - O_t}\,. \tag{1}$$

The vector $t^c_{O_c-O_t}$ corresponds to the translation from the origin O_c of the camera RF to the origin O_t of the target RF, expressed in the camera RF. According to the classical pinhole camera model, the point P^c is projected on the image plane as follows:

$$p = [u_p,\, v_p] = \left[\frac{x^c}{z^c} f + C_x,\, \frac{y^c}{z^c} f + C_y\right], \tag{2}$$

where f denotes the focal length of the camera and (C_x, C_y) the principal point of the image.

We assume that the camera optical axis z_c is always pointing towards the target's COM. During the pose acquisition phase of a RDV, the camera will likely be pointing towards the observed silhouette centroid instead of the target COM. It may be also be the case that, for some operational reasons such as the need for aligning chaser and target docking interfaces, pointing exactly toward the COM cannot be done. However, we demonstrate in Section V-A5 that the proposed pose estimation method is robust to the presence of such pointing errors. Under the assumption of perfect pointing, the degrees of freedom of the problem pass from 6 to 4, since no translation in the camera plane (O_c, x_c, y_c) is present, i.e., $t^c_{O_c-O_t} = [t^c_x\,, t^c_y\,, t^c_z] = [0\,, 0\,, d]$. The objective of the method that we propose is to estimate the quantities d and the Euler angles φ, ϑ, ψ, using a single binary image of the target projected on the image plane.

III. Proposed method

The interest in using global features such as Fourier descriptors or image moments is that these features can be made invariant to translation, to scaling, and most importantly to rotation. If a bi-dimensional shape is described by such invariant features, the value of the features will not depend on the position of the shape centroid (translation invariance), on the shape dimension (scaling invariance), and on the rotation of the shape in the image plane (rotation invariance). With regard to the problem described in Section II, if global invariant features are used do describe the S/C silhouette at a given pose, the value of the features will depend only on the roll and pitch angles $\varphi \in]-\pi, \pi]$ and $\vartheta \in]-\pi/2, \pi/2]$ of the matrix R_{t-c}. In fact, ψ only affects the rotation of the projected shape in the plane image, as seen on Fig.1. On the other hand, the distance d slightly affects the shape of the projected silhouette at a given attitude. However this contribution is negligible if the target's size is considerably smaller than the distance d, as is the case during the pose acquisition phase of a RDV. Thus, it is possible to assume that d affects only the scale. In Section V we analyze the influence of this approximation on the performance of the algorithm.

The principle of the pose estimation algorithm that we propose is the following. During an off-line process, a set of synthetic views of the target, referred to as *training* images in the following, is generated for a sufficiently large number of discrete values of the pairs $(\varphi, \vartheta) \in]-\pi, \pi] \times]-\pi/2, \pi/2]$. When generating the database, the yaw angle ψ and the camera-target distance d remain fixed, with $\psi = \psi_{train}$ set to zero. The choice of distance $d = d_{train}$

can in fact affect the performance of the pose estimation, as we discuss in Section V-A. For a given pair (φ, ϑ), the position of the camera expressed in the target RF is $t^t_{O_t-O_c} = d_{train} \cdot [\sin\vartheta,\ -\cos\vartheta\,\sin\varphi,\ -\cos\vartheta\,\cos\varphi]^T$. Under the assumption that the camera's optical axis is pointing towards the target's COM, the locus of the points $t^t_{O_t-O_c}(d_{train}, \varphi, \vartheta)$ is represented by the sphere of radius d_{train} centered at the target COM. Thus, the relative attitudes used to generate the training images can be assigned by selecting N_w random points on the sphere. In order to avoid oversampling of the polar zone, which would be the result of a uniform sampling of φ and ϑ, the points are assigned using the algorithm suggested in [32]

$$\begin{cases} \varphi &= 2\pi\ \text{rand}_\varphi - \pi \\ \vartheta &= \text{asin}(1 - 2\text{rand}_\vartheta) \end{cases} \tag{3}$$

where rand_φ and rand_ϑ are two independent random variables uniformly distributed in the interval $]0,\ 1]$. Note that although random attitude sampling was used here for benchmarking the different sets of global descriptor, a deterministic sampling may be more adequate for the nominal pose estimation solution. In that case, a spiral scheme could be used to generate uniformly distributed samples on the unit sphere [33].

Once the images corresponding to the N_w viewpoints are generated, the global invariant descriptors are computed for each view. The size of the resulting database is $N_w \times N_f$, with N_f the dimension of the feature vector. At run-time, when the camera acquires a new image of the target (referred to as *test* image in the sequel), the algorithm computes the descriptor vector associated to the resulting view and finds in the database the pair $(\varphi_{meas}, \vartheta_{meas})$ with the closest descriptor vector (minimizing the Euclidean distance), i.e., performs a nearest neighbor search. The two remaining degrees of freedom, i.e., the yaw angle ψ and the relative camera-target distance d, are estimated using geometric moments (GMs), according to the procedure described in the following.

From a mathematical point of view, moments are projections of a function into a polynomial basis [23]. GMs rely on the most intuitive power basis $p_{pq}(x, y) = x^p y^q$. The general definition for a geometric moment of an image is:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y)\ \mathrm{d}x\ \mathrm{d}y\,, \tag{4}$$

where $f(x, y)$ is a function equal to the intensity level of the image in correspondence of the point (x, y). When binary images such as object silhouettes are considered, $f(x, y)$ is equal to 1 inside silhouette and 0 elsewhere. For binary images, the lower order geometric moments are

associated with geometrical properties of the observed silhouette. For example, m_{00} is the area, while $m_{10}/m_{00} = x_c$ and $m_{01}/m_{00} = y_c$ are the x and y coordinates of the silhouette's centroid. Geometric moments can be easily made translation and scale invariant. Translation invariance is obtained by computing the *central* geometric moments μ_{pq}, according to:

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - x_c)^p (y - y_c)^q f(x, y) \, \mathrm{d}x \, \mathrm{d}y \, . \tag{5}$$

Scale invariance is obtained by dividing the central geometric moments by an appropriated power of the area:

$$\nu_{pq} = \frac{\mu_{pq}}{m_{00}^w}, \quad \text{with} \qquad w = \frac{p+q}{2} + 1 \, . \tag{6}$$

From a given image, the yaw angle $\psi \in]-\pi, \pi]$ of R_{c-t} can be retrieved using the second order central moments, as follows. First, the *inclination* $\tilde{\psi}_0 \in]-\pi/2, \pi/2]$ of the major axis of inertia of the silhouette on the image is given by [23]

$$\tilde{\psi}_0 = \frac{1}{2} \mathrm{atan} \left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right) . \tag{7}$$

However, to recover the full in-plane angle of rotation $\psi_0 \in]-\pi, \pi]$ of the silhouette, we still need to determine a specific direction along the axis of inertia, i.e., distinguish between $\psi_0 = \tilde{\psi}_0$ or $\psi_0 = \tilde{\psi}_0 - \pi$. For this, the third order GMs can be used, since they change sign under a rotation of π [34]. Central moments in the image frame rotated by an angle $\tilde{\psi}_0$ are given by

$$\mu'_{pq} = \sum_{l_1=0}^{p} \sum_{l_2=0}^{q} r_{11}^{l_1} r_{21}^{l_2} r_{12}^{p-l_1} r_{11}^{q-l_2} \mu_{l_1+l_2, p+q-(l_1+l_2)}, \tag{8}$$

where $r_{11} = r_{22} = \cos(\tilde{\psi}_0)$, $r_{12} = -r_{21} = \sin(\tilde{\psi}_0)$. Then, as a convention, we define the in-plane rotation ψ_0 to be the direction for which the moment μ'_{30} after rotation by ψ_0 is positive, i.e.,

$$\psi_0 = \begin{cases} \tilde{\psi}_0 & \text{if } \mu'_{30} > 0 \, , \\ \tilde{\psi}_0 - \pi & \text{if } \mu'_{30} < 0 \, . \end{cases} \tag{9}$$

The ambiguity can be resolved only if the silhouette is not rotationally symmetric. In the case of a silhouette having an N-fold rotation symmetry (N-FRS) (i.e., if it repeats itself after rotation around its centroid by $2\pi j/N$, for all $j = 1, ..., N$), N solutions are possible. In such case, only an observation of the target pose on a set of consecutive frames could help solve the ambiguity.

Once $\psi_{0_{test}}$ is determined for the current image, we can deduce the measured yaw angle ψ_{meas} of the camera frame by

$$\psi_{meas} = \psi_{0_{test}} + (\psi_{train} - \psi_{0_{train}}). \tag{10}$$

where $\psi_{0_{train}}$ is the major axis of inertia (precumputed and stored in the database) of the training view that best matches the current view, and $\psi_{train} = 0$ by construction.

Finally, we obtain a measure d_{meas} of the relative camera-target distance along the optical axis from the zeroth order moment $m_{00_{test}}$ of the current silhouette image, namely

$$d_{meas} = \sqrt{m_{00_{train}}/m_{00_{test}}} \cdot d_{train}, \tag{11}$$

where $m_{00_{train}}$ is the zeroth order moment for the best matching training view, which again can be precomputed and stored in the database.

The requirement of perfect pointing introduced in Section II can be relaxed under the hypothesis of a weak perspective model, i.e., when the depth of the object along the line of sight is small compared to the distance from the camera, or when the FOV is relatively small [10]. These are indeed conditions that are met during the pose acquisition phase. In the weak perspective model it is assumed that all points on a 3D object are at the same distance d from the camera without significant errors in the projection with respect to the full pinhole perspective model. Under this approximation, the contribution to the silhouette shape of the relative position and the relative attitude can be decoupled at the cost of an acceptable degradation of the estimation performance. Recalling (2) and exploiting the fact that $d = \sqrt{(t_x^c)^2 + (t_y^c)^2 + (t_z^c)^2} \sim t_z^c$, the components t_x^c and t_y^c can be approximated by

$$\begin{cases} t_x^c &= \dfrac{d}{f}\,(x_{c_{test}} - C_x) \\ t_y^c &= \dfrac{d}{f}\,(y_{c_{test}} - C_y) \end{cases} \tag{12}$$

where $(x_{c_{test}},\, y_{c_{test}})$ are the coordinate of the observed silhouette centroid. The identities (12) can be used to ensure camera pointing even before the target's full pose has been acquired.

By using rotation invariants, we can drastically reduce the dimension of the database that needs to be stored and increase the speed of the search. For example, if the library were built using a uniform discretization of 5 deg for the Euler angles, the database table would contain $N_w = 72 \times 36 \times 72 = 186624$ synthetic views if ψ were included in the search space, but only $N_w = 72 \times 36 = 2592$ synthetic views if rotations invariants are used. In Sections IV-A, IV-B, and IV-C we provide details about the computation of invariant global features using CMs, ZMs, and FDs.

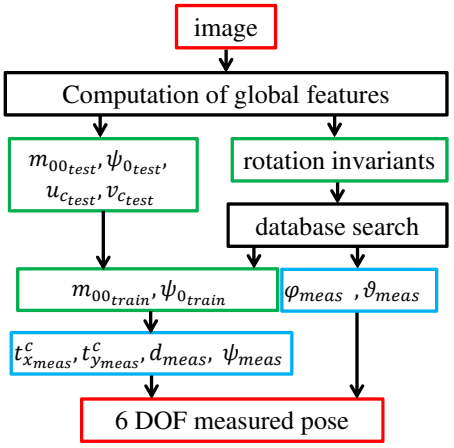


Fig. 2. Structure of the proposed pose estimation algorithm.

IV. Computation of the Invariant Global Features

While designing translation and scale invariants based on moments is generally straightforward, see (5) and (6), rotation invariance is more difficult to obtain. In a seminal paper [35], Hu introduced a set of 7 rotation invariants based on combination of second and third order geometric moments. Unfortunately, Hu's approach cannot be generalized, so that only invariants up to the third order can be derived, limiting the descriptive power of these features. On the other hand, by choosing other polynomial basis functions to compute the moments, as for CMs and ZMs, one can derive general rules for the computation of rotation invariants of up to any order. In particular, a set of invariants up to a given order r must be *independent* and *complete* [23]. A set is independent if none of its elements can be expressed as a function of the other elements. It is complete if any rotation invariant up to the order r can be expressed as a function of the set elements only. A complete set contains both *true invariants* and *pseudo-invariants*. True invariants are invariants that do not change sign under reflection, while pseudo-invariants (also known as *skew-invariants*) do [23]. The capability of pseudo-invariants to detect mirror reflections is very important for the pose estimation problem, since a mirrored image corresponds to a different S/C attitude.

A. Rotation Invariants with Complex Moments

CMs are computed by projecting the image function on the basis $p_{pq}(x, y) = (x+iy)^p(x-iy)^q$, where i is the imaginary unit. The complex moment c_{pq} of order $p+q$ is defined by the following

formula:

$$c_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+iy)^p(x-iy)^q f(x,y)\,\mathrm{d}x\,\mathrm{d}y\,. \tag{13}$$

Translation and scaling invariance can be obtained as for the standard geometric moments, replacing m_{00}, m_{10} and m_{01} in (5) and (6) by c_{00}, $\mathrm{Re}(c_{10})$ and $\mathrm{Im}(c_{10})$ respectively. CMs carry the same amount of information as GMs, but are more convenient to derive an independent and complete set (also referred to as a *basis*) of rotational invariants up to any order, as shown by Flusser [36]. The key points of the approach are recalled here. If the complex moments are expressed in polar coordinates ($x = r\cos\theta$, $y = r\sin\theta$), (13) becomes:

$$\begin{aligned} c_{pq} &= \int_0^{\infty}\int_0^{2\pi}(r\cos\theta + ir\sin\theta)^p(r\cos\theta - ir\sin\theta)^q f(r,\theta)\; r\;\mathrm{d}r\;\mathrm{d}\theta \\ &= \int_0^{\infty}\int_0^{2\pi} r^{p+q+1}\mathrm{e}^{i(p-q)\theta} f(r,\theta)\;\mathrm{d}r\;\mathrm{d}\theta\,. \end{aligned} \tag{14}$$

From (14) we can see that the complex conjugate of a CM satisfies $c_{pq}^* = c_{qp}$ and that a pure rotation of the image around the origin by an angle α changes CMs according to

$$c'_{pq} = \mathrm{e}^{-i(p-q)\alpha} c_{pq}\,, \tag{15}$$

which shows that the magnitude of a CM is a rotation invariant. However CM magnitudes do not generate a complete set. Instead, considering complex moments up to the order $r \geq 2$, we construct the following basis of rotation invariants [36]

$$\mathcal{B}_r = \left\{ \Phi_{pq} := c_{pq}\, c_{q_0p_0}^{p-q} \mid p \geq q \text{ and } p+q \leq r \right\}\,, \tag{16}$$

where p_0 and q_0 that can be arbitrary chosen such that $p_0 + q_0 \leq r$, $p_0 - q_0 = 1$, $q_0 \neq 0$, and $c_{q_0p_0} \neq 0$. Rotation invariance follows from

$$c'_{pq} c'^{p-q}_{q_0p_0} = \mathrm{e}^{-i(p-q)\alpha} c_{pq} \cdot \left(c_{q_0p_0}\mathrm{e}^{-i(q_0-p_0)\alpha}\right)^{p-q} = c_{pq} c_{q_0p_0}^{p-q}\,.$$

The exponents p_0, q_0 are generally chosen as small as possible, e.g., $p_0 = 2$ and $q_0 = 1$, because high order moments are more sensitive to noise [25]. Each basis element Φ_{pq} such that $p \neq q$, except $\Phi_{p_0q_0}$, provides two real-valued invariants, corresponding to the real and imaginary parts of Φ_{pq}. It can be proven that the real part is a true invariant, while the imaginary part is a pseudo-invariant. Hu's 7 invariants can be expressed from the 6 elements of $\mathcal{B}_3$, implying that Hu's set is not independent. Moreover, one element of $\mathcal{B}_3$ cannot be expressed as a function of Hu's invariants, implying that Hu's set is incomplete [23].

Note that if an object has an N-FRS, then all its complex moments with non-integer $(p-q)/N$ are equal to zero. In [37] Flusser provide an extension of (16) for N-FR symmetric objects.

For our pose estimation algorithm, CM invariants up to order 10 (i.e., $N_f = 62$) are computed and stored in the database. The elements $\Phi_{00} = c_{00}$ and $\Phi_{10} = c_{10}c_{12}$ are not included in the set because c_{00} and c_{10} are already used to achieve scaling and translation invariance. Because high-order moments have much higher magnitude than low order moments, the components of the feature vector must be normalized before applying the minimum Euclidean distance criterion. This issue, which is already present with Hu's invariants, has been addressed in the literature using different approaches, e.g., using z-score normalization [17], [20] or variance balancing [18], [29]. Here, we follow the approach suggested in [38] of replacing each feature F by $F_n = \text{sign}(F) \cdot \log(|F|)$, which was found to provide the best recognition capabilities for CM-based invariants.

B. Rotation Invariants with Zernike Moments

ZMs are a family of orthogonal moments, i.e., the corresponding polynomial basis satisfies

$$\iint_{\Omega} p_{pq}(x,y) \cdot p_{jk}(x,y)\, \mathrm{d}x\mathrm{d}y = 0 \tag{17}$$

for all $q \neq j$, $p \neq k$, where Ω is called the *region of orthogonality* and must contain the support of the image f, which must therefore typically be rescaled. For ZMs, Ω is taken to be the unit disk. Given an image expressed in polar coordinates, ZMs are defined by [23]

$$A_{nl} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 V_{nl}^*(r,\theta) f(r,\theta)\, r\, \mathrm{d}r\mathrm{d}\theta \tag{18}$$

where n is a nonnegative integer called the order, $l \in \{-n, -n+2, \ldots, n\}$ is called the repetition (note that the difference $n - |l|$ is always even), and V_{nl} denotes the Zernike polynomials

$$V_{n,l}(r,\theta) = R_{n,l}(r)\mathrm{e}^{il\theta} \tag{19}$$

with radial part

$$R_{n,l}(r) = \sum_{s=0}^{(n-|l|)/2} (-1)^s \frac{(n-s)!}{s!((n+|l|)/2-s)!((n-|l|)/2-s)!} r^{n-2s}\,. \tag{20}$$

The radial functions satisfy $R_{n,-l}(r) = R_{nl}(r)$, so that $A_{n,-l} = A_{nl}^*$ and ZMs with repetition $l = 0$ are real valued moments.

Different methods have been proposed to normalize the image to the unit disk. In [39], each shape is resized so that its zeroth order GM m_{00} is set to a predetermined value, while in [27] a

fixed-dimension bounding box is used instead of the zeroth order moment. In this paper, we simply transform the coordinates of the points belonging to the object's silhouette to normalized central polar coordinates, i.e., $\theta = \text{atan}\left((x - x_c)/(y - y_c)\right)$ and $r = \sqrt{(x - x_c)^2 + (y - y_c)^2}/r_{max}$, with r_{max} the maximum value among the radii of the considered silhouette. This approach, also in [29], ensures that all the points in the object's silhouette are used to compute the moments, and also provides translation invariance. Scaling invariance is obtained by dividing A_{nl} by the zeroth order moment A_{00}, as suggested by [23].

Teague [40] was the first to propose a set of rotation invariants based on ZMs, up to the eighth order, but gave no general rule to derive invariants of higher order. As for CMs, some authors [29], [39], [28] use only the magnitude $|A_{nl}|$, but this provides an incomplete set of invariants because the information carried by the pseudo-invariants is lost. Wallin [41] noted that ZMs, which are complex valued moments, behave as CMs under rotation, so that rotation invariance can be obtained by multiplying ZMs by an appropriate phase-cancellation term, see (15) and the discussion below. As suggested by [23], this term has to be searched among the ZMs with repetition 1, starting from A_{31}. Similarly to CMs, these moments are equal to zero for objects having a rotational symmetry. In this case, the normalization moment should be searched within the ZMs having repetition 2. Given the normalizing moment $A_{n_r l_r}$, the normalized ZM Z_{nl} is

$$Z_{nl} = A_{nl}\mathrm{e}^{-il\phi}, \quad \text{with} \quad \phi = \frac{1}{l_r}\text{atan}\left[\frac{\text{Im}(A_{n_r l_r})}{\text{Re}(A_{n_r l_r})}\right]. \tag{21}$$

Each Z_{nl} leads to two real-valued invariants, except for $Z_{n_r l_r}$ and for the moments with repetition $l = 0$. Z_{00} and Z_{10} are not included in the set. In fact, A_{00} is already used to obtain scaling invariance, while A_{10} is always zero when central coordinates are used. For the pose estimation algorithm, ZM invariants up to the 10^{th} order (i.e., $N_f = 62$) are computed and stored in the database. Unlike GMs and CMs, ZMs values have a smaller dynamic range [23], which simplifies the process of feature matching in the database. Some authors claim that the presence of factorial terms in the radial polynomials increases the computation time needed to compute ZMs, especially for higher order moments, and methods to speed up the computation of the moments are proposed in [42], [43], [44]. However, as discussed in Section V, the optimal performance of the ZM-based pose estimation algorithm is obtained with moments up to the seventh and ninth order. For such relatively low orders, the radial polynomial coefficients can be stored and thus do not need to be computed on-line.

C. Rotation Invariants with Fourier descriptors

FDs provide a representation of the boundary of a two-dimensional shape. Indeed, since a closed curve can be represented by a periodic function of a continuous parameter, it admits a Fourier transform, whose coefficients can be used as global descriptors both for shape recognition and shape retrieval [45]. Different definitions of FDs exist [46], but the most popular starts by defining the complex central coordinate position sequence

$$z(n) = (x(n) - x_c) + i(y(n) - y_c), \quad \text{for } n = 1, \ldots, N_p, \tag{22}$$

with N_p the number of points belonging to the contour and the coordinates of the contour's centroid given by $x_c = \frac{1}{N_p}\sum_{n=1}^{N_p} x(n)$, $y_c = \frac{1}{N_p}\sum_{n=1}^{N_p} y(n)$. The FDs are defined by computing the discrete Fourier transform (DFT) of z using the fast Fourier transform (FFT)

$$Z(\omega) = \sum_{n=1}^{N_p} z(n) \exp\left(-i\frac{2\pi(n-1)}{N_p}\omega\right), \quad \text{for } \omega = 0, \ldots, N_p - 1. \tag{23}$$

Scaling invariance is obtained by dividing the DFT sequence by $|Z(1)|$. Translation invariance is obtained by discarding the coefficient $Z(0)$, which is indeed equal to 0 when complex central coordinates are used to define the position sequence.

The behavior of FDs under rotation is similar to that of CMs and ZMs, see (15). However, the Fourier transform depends also on the starting point used to describe the contour. If this starting point is shifted by m positions, the resulting transform $\tilde{Z}(\omega)$ is

$$\tilde{Z}(\omega) = Z(\omega) \exp\left(\frac{2\pi i}{N_p}\omega m\right). \tag{24}$$

A solution is to use only the magnitude of the transform $|Z(\omega)|$ as descriptor, as in [47], [48], [49], [31], since it is invariant to rotation and independent of the choice of initial point. However the resulting feature set will be incomplete, since every harmonic $Z(\omega)$ provides in fact two invariants, i.e., its magnitude and phase or alternatively real and imaginary parts. As with CMs and ZMs, the imaginary parts of FDs are pseudo-invariants [47]. In [50], the rotation and starting point are determined by a search for the best matching through all the possible shapes, but this procedure increases the computational time of the nearest neighbor search.

In this paper, we propose a method to achieve simultaneously rotation invariance and independence with respect to the choice of initial point. First, using the GMs of the contour, the in-plane rotation angle ψ_0 is computed as described in Section III. By convention, the initial point of the contour is taken to be the intersection of the contour with the half line originating from the figure centroid and with direction ψ_0. In the case of multiple intersections, we select the

farthest one from the centroid. The sequence $X = [x_1 + iy_1,\ ..., x_j + iy_j, ..., x_{N_p} + iy_{N_p}]^T$ is then rearranged starting from the selected initial point and following the contour counterclockwise. Then, the coordinates of the points are rotated by an angle $-\psi_0$, i.e., we let

$$X_r = X\ \mathrm{e}^{-i\psi_0}\,. \tag{25}$$

and compute the DFT of the sequence X_r. This procedure is similar to the approach used in [51] and [22], where the ellipse described by the first harmonic phasor is used instead of the inertia ellipsoid. Note that none of these methods can be applied when ellipse degenerates into a circle, i.e., when the object's silhouette has an N-FRS. Note that curves presenting a N-FRS have zero amplitude harmonics for all indices that are not integral multiples of N [47]. For the pose estimation algorithm, FDs up to the 99^{th} harmonic (i.e., $N_f = 198$) are computed and stored in the database.

V. Application and Performance Analysis

In this section, the performance of the method presented in Section III is characterized for different invariants and conditions. We aim to provide guidelines for choosing the type and number of descriptors (choice of order) among ZMs, CMs and FDs that provide the best trade-off between pose estimation accuracy and computation time. The geometry of the target S/C used

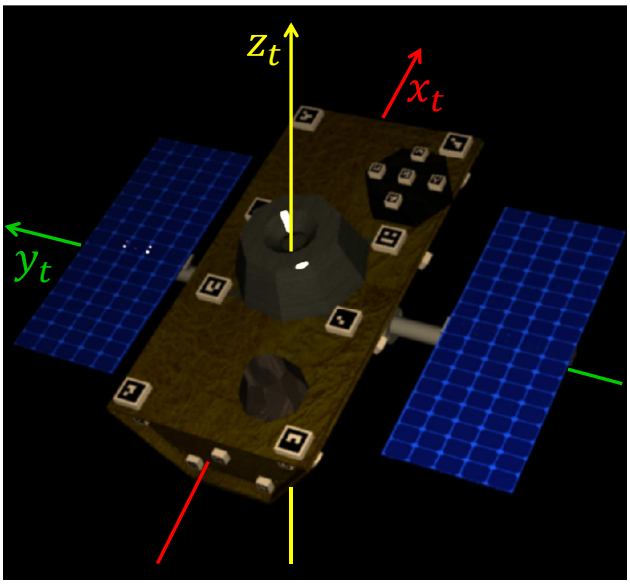
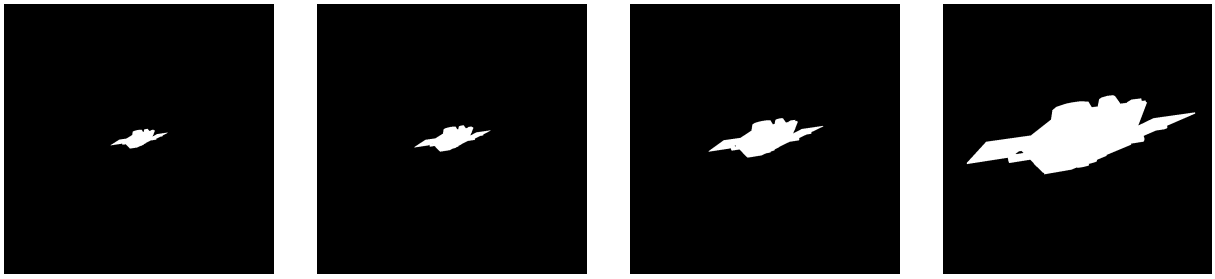


Fig. 3. Geometry and body reference frame of the target spacecraft.

in the simulations is inspired from the structure of the Iridium-NEXT satellites and shown in Fig. 1. The synthetic views are generated with Thales Alenia Space rendering engine SpiCam. The target has a size of $4 \times 3 \times 1$ meters and the views are generated for a camera with a FOV of 30° and image size of 1024×1024 pixels, i.e., on Fig. 1 we have $C_x = C_y = 512$ and $f = C_x / \tan(\text{FOV}/2) = 1911$. The target dimensions, translated in pixel, correspond to a projected size of $191 \times 143 \times 48$ pixels at 40 m, $255 \times 191 \times 64$ pixels at 30 m, $382 \times 287 \times 96$ pixels at 20 m, and $764 \times 573 \times 191$ pixels at 10 m. Figure 4 shows the silhouette of the target at these distances for a given attitude. The main structure of the S/C, composed of the central body and



(a) (b) (c) (d)

Fig. 4. Silhouette of the target satellite at 40 m (Fig.4(a)), 30 m (Fig.4(b)), 20 m (Fig.4(c)), 10 m (Fig.4(d)), for a 1024×1024 pixel camera with FOV of 30°. The attitude of the S/C is $\varphi = 70.41^\circ$, $\vartheta = 66.92^\circ$, $\psi = -29.58^\circ$

the lateral solar arrays, has two symmetry planes, (O_t, x_t, z_t) and (O_t, y_t, z_t), see Fig. 3. Some elements on the central body such as antennas and a docking fixture break the symmetry but are relatively small and visible only for a restricted range of attitudes. The ambiguity of determining the pose of a symmetric body was already noted in previous work on aircraft pose estimation and classification [17], [22], [20]. With the method of [20], for any triplet of Euler angles $[\varphi,\ \vartheta,\ \psi]$, 8 possible solution are possible. Two solutions are due purely to the presence of a symmetry plane, i.e., $[\varphi,\ \vartheta,\ \psi]$ and $[-\varphi,\ -\vartheta,\ \psi - \pi]$. Then, two more solutions, i.e., $[\pi - \varphi,\ -\vartheta,\ \psi]$ and $[\pi + \varphi,\ \vartheta,\ \psi - \pi]$, are due to the impossibility of distinguishing between mirror images using Hu's invariants. Moreover, for all of the 4 solutions described, the method in [20] cannot disambiguate the in-plane rotation between $[\varphi,\ \vartheta,\ \psi]$ and $[\varphi,\ \vartheta,\ \psi - \pi]$. In contrast, by using a complete set of invariants containing also pseudo-invariants, as discussed in Section IV, and by estimating the in-plane rotation as proposed in Section III, we need only to consider the two potential solutions $[\varphi,\ \vartheta,\ \psi]$ and $[-\varphi,\ -\vartheta,\ \psi - \pi]$ for every output of the pose estimation algorithm.

In order to characterize the intrinsic performance of the descriptor without including the problem of distinguish between two symmetric attitudes, the tests described in this Section are done using training images and test images corresponding to attitudes in the semi-sphere where $\varphi \geq 0$. The training database has a size of $N_w = 5000$ and the methods are tested on 2000 images. Both the training and the test attitudes are generated by modifying the sampling scheme (3) to consider only nonnegative values of φ, i.e., with $\varphi = \pi \text{ rand}_\varphi$. For the test attitudes, ψ is also randomly generated such that $\psi = 2\pi \text{ rand}_\psi - \pi$, where rand_ψ is a random variable uniformly distributed in the interval $]0,\ 1]$. Test images are generated with perfect camera pointing (i.e., $t^c_{O_c - O_t} = [0,\ 0,\ d]$), except for the images used in Sec.V-A5, where the algorithm is tested in the presence of camera pointing errors.

Even if the object is symmetric, the number of rotation symmetries of the projected silhouette is always smaller or equal to 1. It is equal to 1 if the projected shape has one axis of reflection [52]. This implies that the invariants proposed in Section IV-A and Section IV-B can be used, except for attitudes corresponding to a camera position close to the sphere poles, i.e., $\vartheta \sim 0$ and $\varphi \sim 0$ or $\varphi \sim \pi$. For these attitudes, the projected S/C shape has a 2-FRS and a different set of invariants should be computed to enable recognition. Nevertheless, according to [23], the order of the normalizing moment used to obtain rotation invariance should be kept as low as possible to improve the performance of the recognition, as higher order moments are more sensitive to noise. In order to avoid degrading the global performance of the algorithm for the sake of improving the recognition of just two isolated orientations, the rotation-normalizing moments are kept equal to c_{12} for CMs and A_{31} for ZMs. Moreover, in correspondence of the polar attitudes, the rule for the determination of the in-plane angle of rotation ψ_0 described in Section III will provide two solutions. However, this ambiguity can be resolved by imposing a continuity constraint between two consecutive pose estimates.

A. Simulation Results

We describe in this section the results of four different simulation experiments. In Paragraph V-A1, our aim is to characterize the degradation of the pose estimation performance when the S/C in the test images is at a distance d_{test} different from the distance d_{train} used to build the database. Pararagraph V-A2 compares the performance of the different global descriptors when the resolution of both test and training images changes. In Paragraph V-A3, the robustness against the resizing of the test images is investigated, while Paragraph V-A4 studies the influence of the

database size N_w on the estimation error. Finally, in Paragraph V-A5, the robustness to camera pointing errors is analyzed.

For any test image, we compute the estimation error in the axis-angle representation using unit quaternions. If $q_{c-t_{true}}$ is the true relative attitude quaternion associated with the camera-target pose in a test image, and $q_{c-t_{meas}}$ the relative attitude quaternion estimated by the algorithm, the error $\delta_{\varphi\vartheta\psi}$ is computed as:

$$\begin{aligned} \delta q &= q^*_{c-t_{true}} \otimes q_{c-t_{meas}} \\ \delta_{\varphi\vartheta\psi} &= 2 \operatorname{atan}\left(\left| \sqrt{\delta q_1^2 + \delta q_2^2 + \delta q_3^2} / \delta q_0 \right| \right) \end{aligned} \tag{26}$$

where quaternions are written $q = q_0 + q_1 i + q_2 j + q_3 k$, with q_0 denoting the scalar part, and q^* denotes the quaternion conjugate $q^* = q_0 - (q_1 i + q_2 j + q_3 k)$. The quantity $\delta_{\varphi\vartheta\psi}$ represents the smallest rotation that aligns the measured quaternion with the true one, and its value is always in the interval $[0, 180°]$. We assume that a test image has found an acceptable match if $\delta_{\varphi\vartheta\psi} < 20°$. Indeed, when using a frame-to-frame tracking algorithm, a pose estimation error in this range can typically be corrected, while outside of this range the matching result can be rejected as outlier, as done in [53]. Finally, we record as performance indices: i) the *accuracy* (also denoted $\% < 20°$), defined here as the percentage of detections with an estimation error lower than $20°$; ii) the *mean* of the error $\delta_{\varphi\vartheta\psi}$ over all the test samples such that $\delta_{\varphi\vartheta\psi} < 20°$; and iii) the *mean* of the measured distance d_{meas} computed using all the test samples.

A baseline test is performed using a database built with $d_{train} = 20$ m and test images taken at the same distance. Table I shows the performance of the CM, ZM, and FD invariants as a function of the moment or harmonic order. The best accuracy for CMs is 90.45%, obtained with invariants of the 5^{th} order, with a mean angular error of $2.67°$. The best accuracy of ZM-based invariants is 96.95%, obtained with moments up to the 9^{th} order, with a mean angular error of $1.74°$. However, little performance improvement is observed beyond the 7^{th} order. The optimal performance of FDs is obtained using coefficients up to the 10^{th} harmonic, with an accuracy of 85.05% and a mean angular error of $3.03°$, and no improvements is observed for higher order harmonics. In general, even when the accuracy of CM, ZM, and FD invariants is comparable (e.g., 4^{th} order CMs, 3^{rd} order ZMs and 10^{th} harmonic FDs), the mean angular error using ZMs is lower. The descriptors show a comparable performance in the estimation of the distance d_{meas}.

1) Effect of a variation of the test distance with a constant training distance: The descriptors based on ZMs, CMs, and FDs are theoretically invariant on a continuous image, but this

TABLE I

PERFORMANCE FOR $d_{train} = 20$ M, $d_{test} = 20$ M.

Complex moment invariants

order:	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}
$\% < 20$ deg	84.15	86.65	**90.45**	89.55	88.40	88.20	86.95	88.50
$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	2.95	2.99	**2.67**	2.75	2.83	2.92	2.92	3.15
d_{meas} *mean* [m]	20.14	20.13	20.15	20.13	20.09	20.08	20.05	**20.02**

Zernike moment invariants

order:	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}
$\% < 20$ deg	85.35	91.85	95.25	95.45	96.40	96.25	**96.95**	96.80
$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	2.12	1.86	1.80	1.74	1.75	**1.73**	1.74	1.76
d_{meas} *mean* [m]	20.07	20.01	20.01	20.01	20.02	**20.01**	20.01	20.01

Fourier descriptors

harmonic:	3^{rd}	4^{th}	5^{th}	7^{th}	9^{th}	10^{th}	20^{th}	99^{th}
$\% < 20$ deg	53.65	74.65	81.50	84.30	84.95	**85.05**	85.05	85.00
$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	4.65	3.34	3.23	3.01	**3.01**	3.03	3.05	3.05
d_{meas} *mean* [m]	20.13	20.21	20.12	20.04	20.03	**20.01**	19.98	19.98

invariance degrades for a digital image due to pixel discretization [54], [31]. Thus, if the descriptors for a test image with the target at distance d_{test} are compared to a database of descriptors computed with the target at a different distance d_{train}, the quality of the matching may be reduced in practice. Moreover, below a certain value of d, it is no more possible to assume that the distance affects only the scale: its contribution to the shape of the projected silhouette may entail an additional degradation of the performance.

To investigate this issue and understand if one can use a database with a single distance to estimate the S/C pose at different distances or if it is necessary to store views covering a wide range of relative distances, we use the same training database built with $d_{train} = 20$ m and the same test attitudes, but perform tests with different values of d_{test}. Results are shown in

TABLE II

EFFECT OF A VARIATION OF THE TEST DISTANCE WITH A CONSTANT TRAINING DISTANCE ($d_{train} = 20$ M)

Complex moment invariants

	order:	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}
$d_{test} = 30$ m	$\% < 20$ deg	73.95	77.55	83.95	**85.80**	83.00	83.15	81.40	82.45
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	5.19	4.94	5.03	**4.90**	5.26	5.21	5.88	5.64
	d_{meas} *mean* [m]	30.45	30.39	30.53	30.36	30.21	30.25	**30.01**	30.02
$d_{test} = 40$ m	$\% < 20$ deg	66.55	71.35	75.65	**78.40**	71.40	73.65	71.00	74.05
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	6.29	6.25	6.46	**6.23**	6.80	6.72	7.35	7.19
	d_{meas} *mean* [m]	40.56	40.59	40.71	40.67	**40.15**	40.24	39.52	39.78
$d_{test} = 10$ m	$\% < 20$ deg	57.20	60.65	64.85	**67.45**	63.30	64.65	64.60	64.60
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	7.93	**7.84**	8.69	8.40	8.92	8.72	8.86	8.82
	d_{meas} *mean* [m]	9.79	9.70	9.75	**9.79**	9.71	9.70	9.72	9.71

Zernike moment invariants

	order:	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}
$d_{test} = 30$ m	$\% < 20$ deg	78.85	86.90	93.00	92.50	93.95	93.20	**94.85**	94.40
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	3.06	2.57	2.55	2.42	2.42	**2.37**	2.41	2.45
	d_{meas} *mean* [m]	30.19	30.11	30.11	**30.11**	30.12	30.12	30.12	30.12
$d_{test} = 40$ m	$\% < 20$ deg	71.40	80.20	89.35	89.65	90.95	89.40	**91.35**	91.25
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	3.90	3.19	3.29	3.02	3.03	**2.96**	3.02	3.11
	d_{meas} *mean* [m]	40.28	40.22	**40.21**	40.22	40.22	40.22	40.23	40.22
$d_{test} = 10$ m	$\% < 20$ deg	62.20	72.90	80.80	84.80	87.00	86.65	**87.70**	87.35
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	5.04	4.64	4.62	**4.11**	4.12	4.15	4.20	4.28
	d_{meas} *mean* [m]	9.87	**9.89**	9.89	9.88	9.87	9.88	9.87	9.87

Fourier descriptors

	harmonic:	3^{rd}	4^{th}	5^{th}	7^{th}	9^{th}	10^{th}	20^{th}	99^{th}
$d_{test} = 30$ m	$\% < 20$ deg	43.80	61.80	69.50	72.10	72.7	72.65	**72.85**	72.75
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	5.69	4.29	4.2	3.70	3.76	**3.75**	3.82	3.82
	d_{meas} *mean* [m]	30.53	30.48	30.32	30.10	30.05	**29.99**	29.92	29.90
$d_{test} = 40$ m	$\% < 20$ deg	34.45	48.80	57.80	60.40	61.10	**61.25**	61.00	60.95
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	6.58	5.10	5.00	**4.42**	4.48	4.50	4.58	4.63
	d_{meas} *mean* [m]	41.08	40.76	40.53	40.16	40.07	**40.01**	39.91	39.90
$d_{test} = 10$ m	$\% < 20$ deg	45.65	80.80	84.60	90.70	**92.25**	91.25	90.25	90.00
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	6.95	5.31	5.14	**4.13**	4.15	4.14	4.22	4.23
	d_{meas} *mean* [m]	9.75	9.77	9.66	9.85	**9.86**	9.84	9.83	9.83

Table II for $d_{test} = 30$ m, $d_{test} = 40$ m and $d_{test} = 10$ m. At 30 m, the maximal accuracy is obtained for CMs up to the 6^{th} order (85.80%), for ZMs up to the 9^{th} order (94.85%), and for FDs up to the 20^{th} harmonic (i.e., 72.85%). Zernike invariants are the most stable, with an accuracy loss of only 2.1% compared to the baseline test. The mean angular error is increased for all the methods. Trends are confirmed when the test distance increases. At 40 m, FDs are the most affected (best accuracy equal to 61.25% for the 10^{th} harmonic), followed by CMs (best accuracy equal to 78.40% with moments up to the 6^{th} order). Zernike invariants confirm their higher stability, with the best performance obtained for the 9^{th} order. The accuracy of 91.35% for ZMs is still higher than the best performance of CM and FD in the baseline test, and the mean angular error of 3.02° is comparable. On the other hand, if the test distance is *decreased* with respect to the training distance, as illustrated in Table II for $d_{test} = 10$ m, the moment-based descriptors (CMs and ZMs) show a substantial performance degradation. The best accuracy for CM invariants (67.45%) is obtained with moments up to the 6^{th} order, with a mean angular error of 8.40°. The degradation of ZM invariant is smaller (accuracy $= 87.70\%$, mean angular error $= 4.20^{\circ}$ for the 9^{th} order), but higher than the degradation incurred at 40 m. This relatively large performance loss can be attributed to the fact that at 10 m the distance starts having a non-negligible contribution to the shape of the projected silhouette. On the other hand, FD shows an opposite trend, with a accuracy of 92.25% for the 9^{th} harmonic, a value which is even higher than the one obtained in the baseline test. The trends are summarized in Fig. 5, which shows as a function of d_{test} the evolution of best performance obtained for each set of descriptors (attained for 6^{th} order CMs, 9^{th} order ZMs, and 10^{th} harmonic for FDs). The measured distance d_{meas} tends to be underestimated when $d_{test} < d_{train}$, and overestimated when $d_{train} < d_{test}$, as a consequence of the effect of d on the shape of the projected silhouette.

2) Effect of a variation in the resolution of both the test and the training images: In order to better isolate in the previous tests the true effect of a mismatch between d_{test} and d_{train} from a potential degradation in performance simply due to a lower silhouette resolution as the distance increases, we performed additional simulation experiments assuming $d_{test} = d_{train}$, with this distance equal to 10 m and 40 m, complementing the baseline test at 20 m. The numerical results are shown in Table III and Fig. 6 displays the performance indices of the 6^{th} order CMs, the 9^{th} order ZMs, and the 10^{th} harmonic FDs as a function of the test distance. The accuracy of CMs and ZMs is only slightly affected by the distance. For ZM invariants up to the 9^{th} order, the accuracy and mean angular error are only slightly affected by the distance, going from

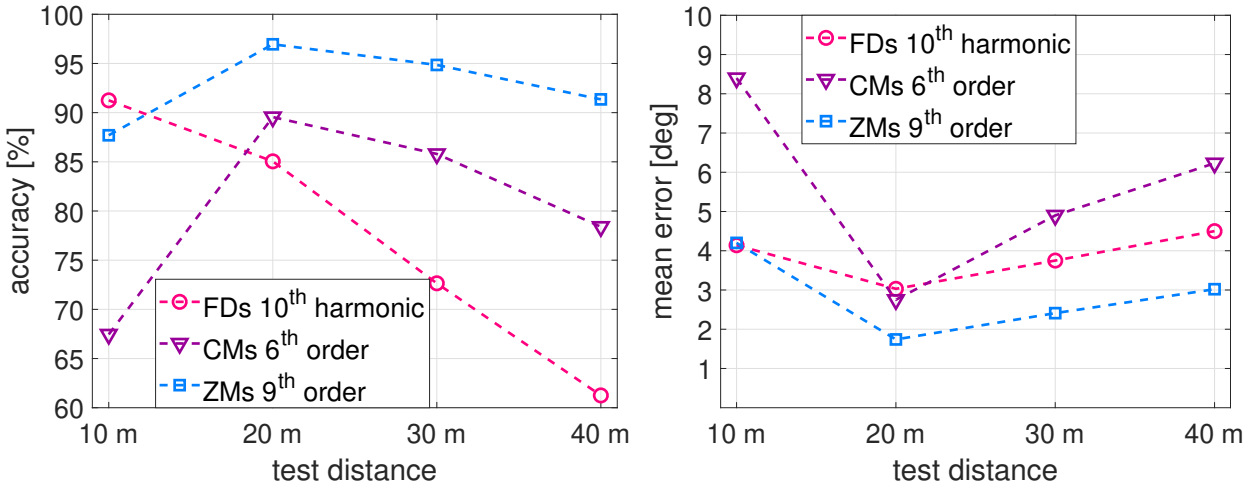


Fig. 5. Effect of a variation of the test distance with a constant training distance $d_{train} = 20$ m.

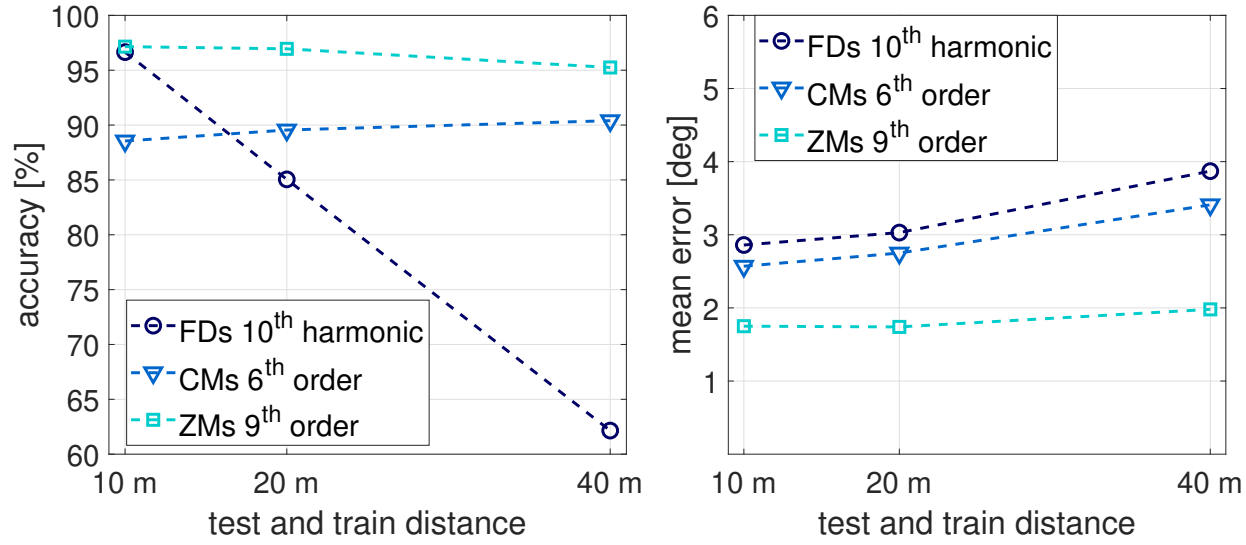


Fig. 6. Effect of a variation in the resolution of both the test and the training images ($d_{test} = d_{train}$).

97.15% and $1.75°$ at 10 m to 95.25% and $1.98°$ at 40 m respectively. On the other hand, the mean angular error of both CMs and FDs degrades more clearly for a lower image resolution. The mean angular error of CM invariants up to the 6^{th} order is equal to $2.57°$ at 10 m and $3.41°$ at 40 m, and the mean angular error of FD invariant up to the 10^{th} harmonic is equal to $2.86°$ at 10 m and $3.87°$ at 40 m. An interesting result is the trend in the accuracy of FDs, which monotonically increases as the image resolution increases, with a maximum of 96.65%

TABLE III

EFFECT OF A VARIATION OF THE RESOLUTION OF BOTH THE TEST AND THE TRAINING IMAGES ($d_{train} = d_{test}$)

Complex moment invariants

	order:	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}
$d_{test} = 40$ m	$\% < 20$ deg	82.00	84.65	**90.75**	90.40	89.30	89.45	88.15	89.15
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	4.07	3.77	3.54	**3.41**	3.54	3.72	3.83	3.92
	d_{meas} *mean* [m]	40.29	40.19	40.19	40.20	40.09	**39.99**	40.02	40.10
$d_{test} = 10$ m	$\% < 20$ deg	82.90	85.05	**89.45**	88.55	87.55	86.60	85.65	85.70
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	2.7	2.69	**2.56**	2.57	2.78	2.91	3.01	3.17
	d_{meas} *mean* [m]	10.08	10.04	10.03	10.04	10.02	10.01	9.99	**10.00**

Zernike moment invariants

	order:	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}
$d_{test} = 40$ m	$\% < 20$ deg	84.35	86.70	93.40	92.95	94.90	94.50	**95.25**	94.75
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	2.41	2.03	2.04	1.97	1.98	1.98	1.98	**1.96**
	d_{meas} *mean* [m]	40.07	40.03	40.04	40.02	40.03	**40.02**	40.03	40.02
$d_{test} = 10$ m	$\% < 20$ deg	85.50	93.05	95.90	96.20	96.80	97.05	97.15	**97.20**
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	1.96	1.87	1.80	1.76	**1.73**	1.74	1.75	1.77
	d_{meas} *mean* [m]	10.03	10.01	10.01	10.00	**10.00**	10.01	10.00	10.01

Fourier descriptors

	harmonic:	3^{rd}	4^{th}	5^{th}	7^{th}	9^{th}	10^{th}	20^{th}	99^{th}
$d_{test} = 40$ m	$\% < 20$ deg	38.45	51.35	59.25	61.85	**62.90**	62.15	62.35	62.25
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	5.65	4.35	4.07	**3.80**	3.93	3.87	3.90	3.88
	d_{meas} *mean* [m]	41.01	40.61	**40.29**	39.69	39.69	39.64	39.60	39.59
$d_{test} = 10$ m	$\% < 20$ deg	62.55	89.20	92.10	96.30	96.50	**96.65**	96.05	95.85
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	4.53	3.30	3.14	**2.79**	2.85	2.86	2.85	2.85
	d_{meas} *mean* [m]	10.05	9.93	9.93	10.00	10.00	**10.00**	10.00	10.00

(10^{th} harmonic) at a distance of 10 m. This performance is comparable to the accuracy of ZM invariants, even though the error obtained with ZMs remains always lower.

3) Effect of a resizing of the test images: The largest contributor to the algorithm latency is the time needed to compute moment invariants (see Sec.V-C). The execution time increases linearly with the number of pixels to be processed, and this may suggest to resize the acquired image

before computing the descriptors. In this paragraph we test the performance of the invariants at $d_{train} = d_{test} = 20$ m, where the test images have been resized from 1024×1024 pixels to 256×256 and 512×512 pixels using OpenCV resize() function [55]. When computing descriptors from the resized images, (11) needs to be corrected to allow computing the correct value of d_{meas}:

$$d_{meas} = \sqrt{m_{00_{train}}/m_{00_{test}}} \cdot d_{train} \cdot \sqrt{n_{pixel_R}/n_{pixel_0}}, \tag{27}$$

where n_{pixel_R} is the total number of pixel in the resized image and n_{pixel_0} is the total number of pixel in the original image. The results are displayed in Table IV. Moment invariants show an accuracy comparable to the baseline test, and a slightly increased angular error. ZM invariants show a higher degradation in the estimation of the distance with respect to CMs. However, this issue can be overcome using the original image to compute m_{00}, and the resized image to compute rotation invariants. The performance of the FD is very poor. However, as discussed in Sec.V-C, the very low time needed to compute FDs suggests that image resizing is unnecessary when working with Fourier invariants.

4) Effect of the database size N_w: Increasing N_w typically leads to a smaller distance between a given target orientation and its nearest neighbor in the database, and hence to a smaller estimation error on average. However, a larger database requires more memory as well as more computation time for matching, although as we discuss below the latter is typically much smaller than the time needed to compute the invariants. The effect of reducing the database size N_w is evaluated for $d_{train} = d_{test} = 20$ m. Table V shows the results for $N_w = 1000$ and $N_w = 3000$, in addition to the baseline scenario with $N_w = 5000$. The estimated average distance is not displayed in the table as no appreciable differences with respect to the baseline test were found. Fig. 7 compares the performance indices as a function of the ZM-based invariants' order for the different values of N_w. As expected, both the accuracy and mean angular error improve with larger values of N_w. However, the marginal improvement also decreases as N_w becomes large, so that above a certain threshold, the performance gains may not be worth the higher storage and computational costs.

5) Effect of a pointing error: While the database is constructed by centering the image at the target's COM, during the pose acquisition phase in a real RDV the camera will likely point at the silhouette's centroid, causing a small pointing error. In some scenario, it may be even impossible to keep the target in the center of the FOV, causing a larger pointing error.

TABLE IV

EFFECT OF A RESIZING OF THE TEST IMAGE. ($d_{train} = d_{test} = 20$ M)

Complex moment invariants

	order:	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}
256×256	$\% < 20$ deg	78.30	82.75	87.40	**87.55**	86.25	86.25	85.40	86
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	3.85	3.69	**3.49**	3.52	3.58	3.63	3.74	3.9
	d_{meas} *mean* [m]	20.02	19.99	**19.99**	19.98	19.96	19.92	19.91	19.86
512×512	$\% < 20$ deg	81.9	84.70	**89.90**	88.90	87.00	87.10	86.30	87.30
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	3.53	3.43	3.16	**3.15**	3.30	3.41	3.52	3.64
	d_{meas} *mean* [m]	19.98	19.97	19.99	**20.00**	19.96	19.93	19.84	19.86

Zernike moment invariants

	order:	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}
256×256	$\% < 20$ deg	78.55	89.65	94.20	94.30	95.25	95.55	**95.95**	95.90
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	3.52	2.84	2.70	2.49	2.46	2.42	**2.38**	2.41
	d_{meas} *mean* [m]	**20.02**	19.92	19.91	19.91	19.90	19.90	19.90	19.89
512×512	$\% < 20$ deg	83.05	91.40	94.55	94.60	95.60	95.55	**96.55**	96.35
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	2.66	2.20	2.11	1.98	1.95	**1.95**	1.96	1.98
	d_{meas} *mean* [m]	**19.94**	19.89	19.88	19.87	19.87	19.87	19.87	19.87

Fourier descriptors

	harmonic:	3^{rd}	4^{th}	5^{th}	7^{th}	9^{th}	10^{th}	20^{th}	99^{th}
256×256	$\% < 20$ deg	25.50	40.35	45.45	48.40	48.75	48.65	48.75	**48.80**
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	7.44	5.98	5.74	**5.30**	5.39	5.35	5.37	5.40
	d_{meas} *mean* [m]	20.39	20.12	**20.10**	19.55	19.52	19.47	19.40	19.41
512×512	$\% < 20$ deg	37.4	55.15	60.70	63.35	**63.80**	63.65	63.60	63.70
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	5.88	4.46	4.26	**4.01**	4.03	4.05	4.15	4.16
	d_{meas} *mean* [m]	20.24	20.12	**20.04**	19.71	19.68	19.64	19.61	19.60

The presence of this error affects not only the position of the image's centroid (which does not influence the invariants), but also the projected shape of the S/C, because of the camera's perspective projection. It is therefore important to understand whether the presence of such pointing errors affects the algorithm. To do this, we added angular offsets to the camera RF around the x_c and y_c axis and tested the algorithm at a distance of 20 m, using the database

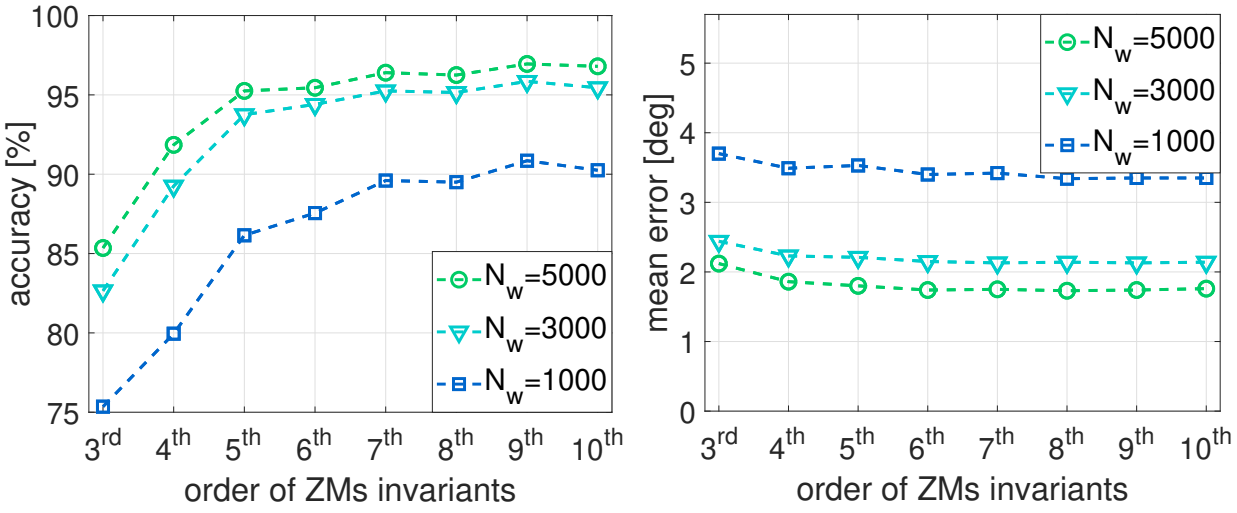
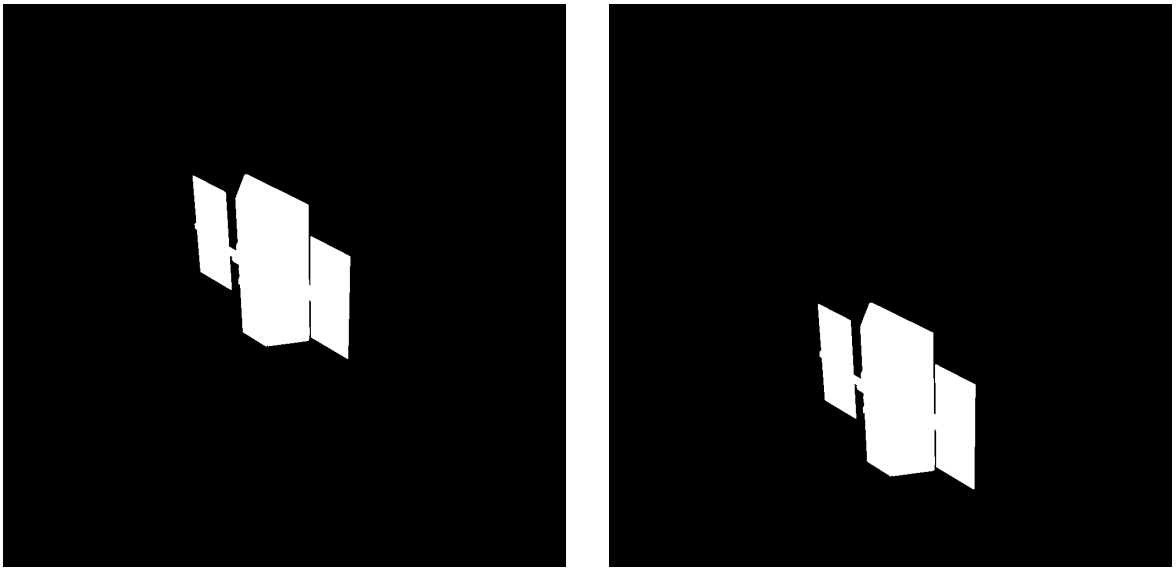


Fig. 7. Performance of the ZM invariants for different database sizes N_w.



(a) (b)

Fig. 8. Silhouette of the target satellite at 20 m, at a relative attitude of $\vartheta = -35.02°$, $\psi = 87.99°$. In Fig.8(a) the pointing error is $\Delta\gamma = 1.43°$, in Fig.8(b) it is $\Delta\gamma = 5.85°$.

built with $d_{train} = 20$ m. Two sets of 2000 images were generated. In the first one, each image is generated adding a total pointing error $\Delta\gamma$ such that $\Delta\gamma \in [1.42°, 2°]$ (see Fig.8(a)). In the second set, the error is $\Delta\gamma \in [4.25°, 6°]$ (see Fig.8(b)). For these tests we collect also the performance on the estimation of the relative position vector $t^c_{O_c - O_{tmeas}}$. In particular, the *mean*

TABLE V

EFFECT OF DATABASE SIZE N_w. WITH $d_{train} = d_{test} = 20$ M.

Complex moment invariants

	order:	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}
$N_w = 1000$ m	$\% < 20$ deg	71.80	74.10	**77.95**	77.30	76.70	75.10	74.70	75.95
	mean [deg]	**4.82**	4.94	4.92	5.05	5.27	5.26	5.17	5.22
$N_w = 3000$ m	$\% < 20$ deg	80.30	82.95	**86.85**	86.80	85.65	85.75	84.10	84.75
	mean [deg]	3.35	3.44	**3.23**	3.36	3.51	3.54	3.50	3.68

Zernike moment invariants

	order:	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}
$N_w = 1000$ m	$\% < 20$ deg	75.35	79.95	86.15	87.55	89.60	89.50	**90.85**	90.25
	mean [deg]	3.70	3.49	3.53	3.40	3.42	**3.34**	3.35	3.35
$N_w = 3000$ m	$\% < 20$ deg	82.65	89.25	93.75	94.40	95.25	95.15	**95.85**	95.45
	mean [deg]	2.44	2.23	2.21	2.15	2.13	2.14	**2.13**	2.14

Fourier descriptors

	harmonic:	3^{rd}	4^{th}	5^{th}	7^{th}	9^{th}	10^{th}	20^{th}	99^{th}
$N_w = 1000$ m	$\% < 20$ deg	49.50	68.70	75.30	77.90	**79.20**	78.95	79.10	79.05
	mean [deg]	5.36	4.56	4.46	**4.29**	4.37	4.41	4.48	4.49
$N_w = 3000$ m	$\% < 20$ deg	53.15	73.20	80.55	82.75	83.15	83.35	**83.40**	83.20
	mean [deg]	4.86	3.68	3.56	**3.33**	3.34	3.38	3.42	3.42

of the absolute errors Δt_x^c, Δt_y^c, Δt_z^c, is shown. The results are displayed in Table VI and show that all the methods are relatively robust to pointing errors. The accuracy of the invariants remains almost the same as for the baseline test, while the mean angular errors increase slightly. It is interesting to see how the attitude error grows almost linearly with the pointing error $\Delta\gamma$. ZM based invariants show lower angular errors and position errors than the other sets of invariants.

B. Discussion of the results

From the tests described in this section, the following important conclusions can be drawn.

TABLE VI

EFFECT OF A POINTING ERROR. $d_{train} = d_{test} = 20$ M

Complex moment invariants

	order:	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}
$\Delta\gamma \in [1.41°, 2°]$	$\% < 20$ deg	83.00	86.35	**90.40**	89.80	88.85	88.50	87.65	88.50
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	3.57	3.55	3.36	**3.43**	3.53	3.62	3.65	3.80
	Δt_x^c *mean* [cm]	4.75	4.67	**4.63**	4.65	4.71	4.65	4.74	4.74
	Δt_y^c *mean* [cm]	4.89	4.83	4.80	4.78	**4.69**	4.75	4.88	4.81
	Δt_z^c *mean* [m]	0.605	0.537	**0.499**	0.521	0.539	0.563	0.638	0.624
$\Delta\gamma \in [4.25°, 6°]$	$\% < 20$ deg	80.30	84.70	**88.05**	87.65	85.80	85.95	85.70	86.25
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	6.08	6.01	5.84	**5.83**	5.89	5.93	6.17	6.26
	Δt_x^c *mean* [cm]	6.70	6.20	6.32	**6.05**	6.36	6.20	6.64	6.62
	Δt_y^c *mean* [cm]	6.82	6.18	6.11	**5.96**	6.42	6.54	6.59	6.53
	Δt_z^c *mean* [m]	0.701	0.577	0.587	**0.553**	0.654	0.641	0.691	0.690

Zernike moment invariants

	order:	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}
$\Delta\gamma \in [1.41°, 2°]$	$\% < 20$ deg	85.40	91.85	95.70	95.70	96.65	96.85	**97.35**	97.15
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	2.77	2.57	2.54	2.47	**2.45**	2.47	2.47	2.47
	Δt_x^c *mean* [cm]	4.49	4.39	4.39	4.38	**4.37**	4.37	4.38	4.37
	Δt_y^c *mean* [cm]	4.53	4.46	4.46	4.46	**4.45**	4.46	4.46	4.45
	Δt_z^c *mean* [m]	0.292	0.162	0.156	0.156	**0.156**	0.157	0.157	0.157
$\Delta\gamma \in [4.25°, 6°]$	$\% < 20$ deg	84.25	90.70	94.90	94.60	95.90	96.05	**97.10**	96.65
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	5.47	5.29	5.27	5.24	**5.20**	5.21	5.21	5.22
	Δt_x^c *mean* [cm]	5.07	4.54	4.56	4.53	4.52	**4.50**	4.50	4.50
	Δt_y^c *mean* [cm]	5.16	4.71	4.68	4.66	4.65	**4.63**	4.63	4.63
	Δt_z^c *mean* [m]	0.318	0.176	0.176	0.169	0.168	**0.166**	0.167	0.168

Fourier descriptors

	harmonic:	3^{rd}	4^{th}	5^{th}	7^{th}	9^{th}	10^{th}	20^{th}	99^{th}
$\Delta\gamma \in [1.41°, 2°]$	$\% < 20$ deg	53.45	74.15	80.65	84.15	**84.80**	84.60	84.55	84.50
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	4.99	3.88	3.62	3.48	**3.48**	3.49	3.51	3.51
	Δt_x^c *mean* [cm]	6.84	5.67	5.53	4.72	**4.70**	4.71	4.72	4.72
	Δt_y^c *mean* [cm]	6.56	5.63	5.45	**4.86**	4.88	4.89	4.89	4.89
	Δt_z^c *mean* [m]	2.211	1.235	1.088	**0.586**	0.593	0.599	0.610	0.613
$\Delta\gamma \in [4.25°, 6°]$	$\% < 20$ deg	53.05	75.00	81.00	84.10	84.45	**84.60**	84.45	84.45
	$\delta_{\varphi\vartheta\psi}$ *mean* [deg]	6.81	6.05	5.78	**5.74**	5.75	5.77	5.79	5.77
	Δt_x^c *mean* [cm]	13.26	8.66	8.04	5.82	**5.81**	5.84	5.93	5.91
	Δt_y^c *mean* [cm]	14.08	9.10	8.69	**6.27**	6.28	6.31	6.29	6.28
	Δt_z^c *mean* [m]	2.189	1.166	1.038	**0.562**	0.566	0.571	0.584	0.585

1) If the target is close, the shape of its projected silhouette can change significantly with the distance, resulting in a significant decrease in performance for moment-based invariants. This phenomenon depends on the camera sensor properties and on the target geometry. In our experiments, it starts occurring for a distance of about 10 m, i.e., about two-hand-a-half times the maximum dimension of the target. FDs, on the other hand, suffer less from this issue.
2) The accuracy of moment-based invariants is only mildly affected by the image resolution, see Fig. 6, and by the image resizing, see Par. V-A3. Thus, images could be sub-sampled before computing moments in order to decrease the computation time of the algorithm. This results in a small performance loss, which might be acceptable if the algorithm is only used to detect the divergence of a classical iterative tracking algorithm.
3) ZM invariants up to the 4^{th} order (i.e., 13 features) always perform better than CM and FD invariants of any order or harmonic. The accuracy for FDs is comparable to that of ZMs only for high resolution images.
4) The accuracy of FDs is highly affected by the test distance, and the accuracy always increases as the test distance decreases, regardless of the training distance. This can be explained as follows. The majority of the spectral content of the projected S/C silhouette, which is a relatively simple shape, is contained in the first 10 harmonics. For a low resolution image, the rasterization effect shifts some of the spectral content of the S/C shape to be to higher frequencies, resulting in loss of information. On the other hand, for high resolution images, the spectral content is correctly distributed in the first harmonics and the matching accuracy is enhanced. This feature is also very interesting because for short distances the computation time for the silhouette's moments increases due to the large number of pixels to be processed. The computation of FDs on the other hand needs less time, since only contour points have to be processed. Thus, it could be useful to switch from moment-based descriptors to FDs as the test distance becomes sufficiently small.
5) The distance is always correctly recovered using (11), even when $\delta_{\varphi\vartheta\psi}$ is higher than $20°$. Hence, even in the case of an incorrect matching, the best match is an image having a "mass" distribution similar to the one of the S/C silhouette. This provides a method to estimate the camera-target distance d using monocular vision, even when it is not possible to estimate the attitude correctly.

It should be noted that all the tests have been carried out using perfect binary images, and that a degradation should be expected when using real images of the target spacecraft. When using

real images, the extracted binary silhouettes are very likely to be affected by segmentation errors. However, the design or choice of robust background subtraction and segmentation algorithms are out of the scope of this study.

C. Computation Time and Memory Requirements

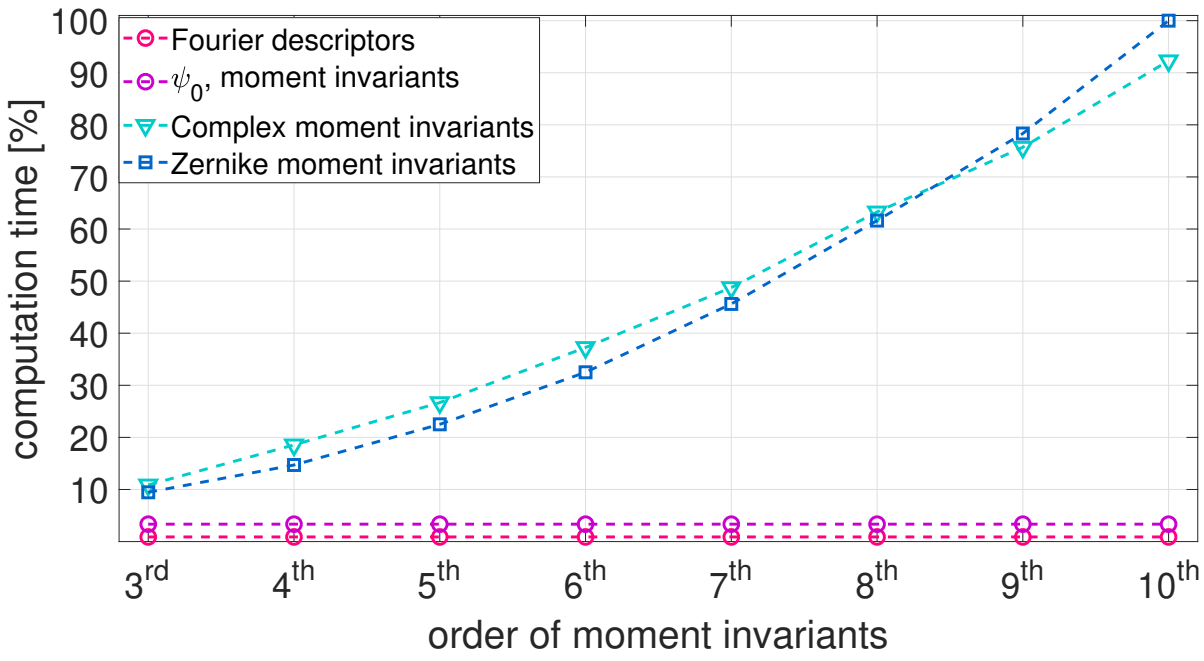


Fig. 9. Computation time normalized with respect to the 10^{th} order ZM invariant computation time

Fig.9 shows the evolution of the computation time needed to compute the descriptors, as a function of the invariants' order, averaged over the images of the test set at 20 m. The times are normalized with respect to the time needed to compute ZM invariants up to the 10^{th} order. Using MATLAB on a 2.70 GHz Intel Core i7 processor, this time is equal to 649.95 milliseconds. Note however that this absolute computation time is mostly indicative, since the implementation could be optimized and would be done in a statically compiled programming language or directly in hardware on a real system, which can lead to significant improvements.

For ZMs and CMs, the total computation time of the algorithm is the sum of the time needed to compute the n^{th} order invariants and the time needed to compute the in-plane rotation angle ψ_0. For FDs, the computation time includes the time needed to extract the edges from the silhouette's

image (using the OpenCV function findContours() [56]), as well as the time to compute the in-plane rotation. In fact, for FDs, ψ_0 must be calculated before computing the Fourier transform. The FFT computes all the harmonics up to the size of the input sequence simultaneously, hence the computation time does not depend on the invariants' order.

We see that the computation time for CMs and ZMs is comparable, while FDs are much faster to compute than moment-based descriptors. Note that the computation of the 7^{th} order ZM invariants requires only half of the time needed to compute the 10^{th} order ones, and that in every test, these invariants have shown an accuracy very close to the optimal one (which was always obtained for the 9^{th} or 10^{th} order) as well as similar mean angular errors. Hence, 7^{th} order ZM invariant offer a good compromise between computational cost and estimation performance. CM invariants offer no particular advantage, as they are always outperformed by ZM invariants and take essentially the same amount of time to compute. Finally, the size of the database affects only moderately the overall computation time of the algorithm, since the time necessary for the nearest neighbor search is largely dominated by the descriptor computation time. For example, with a database of size $N_w = 5000$, matching requires less than 0.1% of the time needed to compute the 10^{th} order ZM invariants. Thus, the choice of the value of N_w should be driven only by the memory available to store the database and by the precision required for the algorithm. Note again that since the size of the database grows exponentially with the number of parameters to discretize, this analysis relies crucially on the fact that the distance and yaw angles are not included in the matching process, thanks to the invariance properties of the descriptors. The memory needed for the descriptor database storage can be computed by multiplying the value of 8 bytes (i.e., the dimension of a *double* type variable) by the total number of *doubles* in the database, which is given by $N_w \times (N_f + 4)$ (i.e., the number of the rotation invariant N_f, plus $m_{00_{train}}$, $\psi_{0_{train}}$, φ_{train}, and ϑ_{train}). A database of ZMs invariants up to the 9^{th} order (i.e., $N_f = 53$) with $N_w = 5000$ has a size of 2.28 Mbytes. A navigation solution using two databases of ZMs invariants up to the 7^{th} order (i.e., $N_f = 34$) computed respectively for $d_{train} = 40$ m and $d_{train} = 20$ m, and a database of FDs up to the 10^{th} harmonic (i.e., $N_f = 18$) computed at $d_{train} = 10$ m, requires a size of 3.93 MBytes if $N_w = 5000$, and 2.353 MBytes if $N_w = 3000$. These requirements are compatible with the resources available on typical space qualified avionics.

VI. Conclusion

This article proposes a template matching method to estimate the pose of a non-cooperative target during space rendezvous from a single binary image capturing the target's silhouette. The method is suitable for initial pose acquisition and for detecting faults and deviations in other on-board trackers, and represents a novel approach for the pose estimation of a spacecraft in a rendezvous. Three types of global descriptors, based on complex moments, Zernike moments, and Fourier descriptors, are introduced and compared in order to match the silhouette in a database of pose-dependent feature vectors generated offline. By exploiting the scale and rotation invariance of these descriptors, the approach requires discretizing only two pose angles to construct the database, leading to fast computation times appropriate for real-time implementations. Our performance analysis on simulated images shows that Zernike moment invariants provide the highest accuracy and robustness in off-nominal conditions. Fourier descriptors show comparable performance with a much lower computational cost, but only on for high resolution images or short target distances. This suggests that these two types of descriptors are complementary and could be used in combination. Future work will focus on the post-processing of the algorithm outputs to detect outliers and discriminate between symmetric attitudes.

References

[1] CONFERS, "Satellite servicing safety framework technical and operational guidance document," https://www.satelliteconfers.org/publications/, 2018, online; [retrieved 06 December 2019].

[2] J. M. Kelsey, J. Byrne, M. Cosgrove, S. Seereeram, and R. K. Mehra, "Vision-based relative pose estimation for autonomous rendezvous and docking," in *2006 IEEE aerospace conference*. IEEE, 2006, pp. 20–pp.

[3] A. Sonnenburg, M. Tkocz, and K. Janschek, "Ekf-slam based approach for spacecraft rendezvous navigation with unknown target spacecraft," *IFAC Proceedings Volumes*, vol. 43, no. 15, pp. 339–344, 2010.

[4] M. Dor and P. Tsiotras, "Orb-slam applied to spacecraft non-cooperative rendezvous," in *2018 Space Flight Mechanics Meeting*, 2018, p. 1963.

[5] T. Lei, X.-F. Liu, G.-P. Cai, Y.-M. Liu, and P. Liu, "Pose estimation of a noncooperative target based on monocular visual slam," *International Journal of Aerospace Engineering*, vol. 2019, 2019.

[6] V. Lepetit, P. Fua *et al.*, "Monocular model-based 3d tracking of rigid objects: A survey," *Foundations and Trends® in Computer Graphics and Vision*, vol. 1, no. 1, pp. 1–89, 2005.

[7] S. Sharma, J. Ventura, and S. D'Amico, "Robust model-based monocular pose initialization for noncooperative spacecraft rendezvous," *Journal of Spacecraft and Rockets*, vol. 55, no. 6, pp. 1414–1429, 2018.

[8] V. Pesce, R. Opromolla, S. Sarno, M. Lavagna, and M. Grassi, "Autonomous relative navigation around uncooperative spacecraft based on a single camera," *Aerospace Science and Technology*, vol. 84, pp. 1070–1080, 2019.

[9] D. Rondao and N. Aouf, "Multi-view monocular pose estimation for spacecraft relative navigation," in *2018 AIAA Guidance, Navigation, and Control Conference*, 2018, p. 2100.

[10] A. Petit, E. Marchand, and K. Kanani, "A robust model-based tracker combining geometrical and color edge information," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 3719–3724.

[11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 886–893.

[12] S. Hinterstoisser, V. Lepetit, S. Ilic, P. Fua, and N. Navab, "Dominant orientation templates for real-time detection of texture-less objects," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 2257–2264.

[13] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, and H. C. Wolf, "Parametric correspondence and chamfer matching: Two new techniques for image matching," in *Proceedings: Image Understanding Workshop*. Science Applications, Inc Arlington, VA, 1977, pp. 21–27.

[14] S. Sharma and S. D'Amico, "Neural network-based pose estimation for noncooperative spacecraft rendezvous," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 1–1, 2020.

[15] S. Sharma, C. Beierle, and S. D'Amico, "Pose estimation for non-cooperative spacecraft rendezvous using convolutional neural networks," in *2018 IEEE Aerospace Conference*. IEEE, 2018, pp. 1–12.

[16] M. Kisantal, S. Sharma, T. H. Park, D. Izzo, M. Märtens, and S. D'Amico, "Satellite pose estimation challenge: Dataset, competition design and results," *IEEE Transactions on Aerospace and Electronic Systems*, 2020.

[17] S. A. Dudani, K. J. Breeding, and R. B. McGhee, "Aircraft identification by moment invariants," *IEEE transactions on computers*, vol. 100, no. 1, pp. 39–46, 1977.

[18] A. P. Reeves, R. J. Prokop, S. E. Andrews, and F. P. Kuhl, "Three-dimensional shape analysis using moments and Fourier descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 6, pp. 937–943, 1988.

[19] T. Glais and A. Ayoun, "Image-based air target identification," in *Applications of Digital Image Processing XVII*, vol. 2298. International Society for Optics and Photonics, 1994, pp. 540–551.

[20] M. G. Breuers, "Image-based aircraft pose estimation using moment invariants," in *Automatic Target Recognition IX*, vol. 3718. International Society for Optics and Photonics, 1999, pp. 294–304.

[21] T. P. Wallace and P. A. Wintz, "An efficient three-dimensional aircraft recognition algorithm using normalized Fourier descriptors," *Computer Graphics and Image Processing*, vol. 13, no. 2, pp. 99–126, 1980.

[22] Z. Chen and S.-Y. Ho, "Computer vision for robust 3D aircraft recognition with fast library search," *Pattern recognition*, vol. 24, no. 5, pp. 375–390, 1991.

[23] J. Flusser, B. Zitova, and T. Suk, *Moments and moment invariants in pattern recognition*. John Wiley & Sons, 2009, pp 7,13-30,186-197.

[24] Y. S. Abu-Mostafa and D. Psaltis, "Recognitive aspects of moment invariants," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 698–706, 1984.

[25] C.-H. Teh and R. T. Chin, "On image analysis by the methods of moments," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 10, no. 4, pp. 496–513, 1988.

[26] S. Gishkori and B. Mulgrew, "Pseudo-zernike moments based sparse representations for sar image classification," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 2, pp. 1037–1044, 2019.

[27] F. Gandía and A. G. Casas, "Fast spacecraft pose estimation based on Zernike moments," in *Proceedings of the 7th International Symposium on Artificial Intelligence, Robotics and Automation in Space, Nara, Japan*, 2003, pp. 19–23.

[28] K.-y. Chang and J. Ghosh, "Three-dimensional model-based object recognition and pose estimation using probabilistic principal surfaces," in *Applications of Artificial Neural Networks in Image Processing V*, vol. 3962. International Society for Optics and Photonics, 2000, pp. 192–203.

[29] A. Khotanzad and J.-H. Liou, "Recognition and pose estimation of unoccluded three-dimensional objects from a two-dimensional perspective view by banks of neural networks," *IEEE Transactions on Neural networks*, vol. 7, no. 4, pp. 897–906, 1996.

[30] T. Fu and X. Sun, "The relative pose estimation of aircraft based on contour model," in *International Conference on Optical and Photonics Engineering (icOPEN 2016)*, vol. 10250. International Society for Optics and Photonics, 2017, p. 102502T.

[31] A. L. Barczak, A. Gilman, N. H. Reyes, and T. Susnjak, "Analysis of feature invariance and discrimination for hand images: Fourier descriptors versus moment invariants," in *International Conference Image and Vision Computing New Zealand IVCNZ2011*, 2011.

[32] J. J. Kuffner, "Effective sampling and distance metrics for 3d rigid body path planning," in *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*, vol. 4. IEEE, 2004, pp. 3993–3998.

[33] C. G. Koay, "Analytically exact spiral scheme for generating uniformly distributed points on the unit sphere," *Journal of computational science*, vol. 2, no. 1, pp. 88–91, 2011.

[34] O. Tahri, "Application des moments a l'asservissement visuel et au calcul de pose," *These de doctorat, Université de Rennes*, vol. 1, 2004.

[35] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE transactions on information theory*, vol. 8, no. 2, pp. 179–187, 1962.

[36] J. Flusser, "On the independence of rotation moment invariants," *Pattern recognition*, vol. 33, no. 9, pp. 1405–1410, 2000.

[37] J. Flusser and T. Suk, "Rotation moment invariants for recognition of symmetric objects," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3784–3790, 2006.

[38] S. Mallick, "Shape matching using Hu moments," https://www.learnopencv.com/shape-matching-using-hu-moments-c-python/, 2018, online; [retrieved 23 April 2020].

[39] A. Khotanzad and Y. H. Hong, "Invariant image recognition by Zernike moments," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 12, no. 5, pp. 489–497, 1990.

[40] M. R. Teague, "Image analysis via the general theory of moments," *JOSA*, vol. 70, no. 8, pp. 920–930, 1980.

[41] Å. Wallin and O. Kubler, "Complete sets of complex Zernike moment invariants and the role of the pseudoinvariants," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 11, pp. 1106–1110, 1995.

[42] A. Prata and W. Rusch, "Algorithm for computation of Zernike polynomials expansion coefficients," *Applied Optics*, vol. 28, no. 4, pp. 749–754, 1989.

[43] S.-K. Hwang and W.-Y. Kim, "A novel approach to the fast computation of Zernike moments," *Pattern Recognition*, vol. 39, no. 11, pp. 2065–2076, 2006.

[44] C.-Y. Wee, R. Paramesran, and F. Takeda, "New computational methods for full and subset Zernike moments," *Information Sciences*, vol. 159, no. 3-4, pp. 203–220, 2004.

[45] K. Arbter, W. E. Snyder, H. Burkhardt, and G. Hirzinger, "Application of affine-invariant Fourier descriptors to recognition of 3-D objects," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 12, no. 7, pp. 640–647, 1990.

[46] D. Zhang, G. Lu *et al.*, "A comparative study of Fourier descriptors for shape representation and retrieval," in *Proc. 5th Asian Conference on Computer Vision*. Citeseer, 2002, p. 35.

[47] C. T. Zahn and R. Z. Roskies, "Fourier descriptors for plane closed curves," *IEEE Transactions on computers*, vol. 100, no. 3, pp. 269–281, 1972.

[48] Q. Chen, E. Petriu, and X. Yang, "A comparative study of Fourier descriptors and Hu's seven moment invariants for image recognition," in *Canadian conference on electrical and computer engineering 2004 (IEEE Cat. No. 04CH37513)*, vol. 1. IEEE, 2004, pp. 103–106.

[49] S. Conseil, S. Bourennane, and L. Martin, "Comparison of Fourier descriptors and Hu moments for hand posture recognition," in *2007 15th European Signal Processing Conference*. IEEE, 2007, pp. 1960–1964.

[50] E. Persoon and K.-S. Fu, "Shape discrimination using Fourier descriptors," *IEEE Transactions on systems, man, and cybernetics*, vol. 7, no. 3, pp. 170–179, 1977.

[51] F. P. Kuhl and C. R. Giardina, "Elliptic Fourier features of a closed contour," *Computer graphics and image processing*, vol. 18, no. 3, pp. 236–258, 1982.

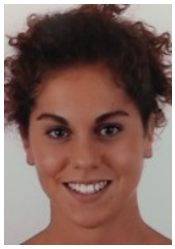
[52] D. Shen, H. H.-S. Ip, and E. K. Teoh, "A novel theorem on symmetries of 2D images," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, vol. 3. IEEE, 2000, pp. 1002–1005.

[53] A. Comellini, F. Maye, D. Casu, E. Zenou, V. Dubanchet, and C. Espinosa, "Robust navigation solution for vision based autonomous rendezvous," *2021 IEEE Aerospace Conference*, pp. 1–14, 2021.

[54] Z. Huang and J. Leng, "Analysis of Hu's moment invariants on image scaling and rotation," in *2010 2nd International Conference on Computer Engineering and Technology*, vol. 7. IEEE, 2010, pp. V7–476.

[55] OpenCV development team, "Geometric transformations of images," https://docs.opencv.org/master/da/d6e/tutorial_py_geometric_transformatio 2019, online; [retrieved 24 November 2020].

[56] ——, "Shape matching using hu moments," https://docs.opencv.org/2.4/doc/tutorials/imgproc/shapedescriptors/find _contours/find_contours.html, 2019, online; [retrieved 28 April 2020].



Anthea Comellini received the M.Sc double-degree in space engineering from Politecnico di Milano, Italy, and ISAE-SUPAERO, France, in 2017, and the M.Sc. in automation and signal processing from Paris-Saclay University, France, in 2018. She is currently pursuing the Ph.D. degree in space engineering with ISAE-SUPAERO and Thales Alenia Space, France. Her research interests include vision-based navigation and GNC for space rendezvous.

Jerome Le Ny (S'05-M'09-SM'16) received the Engineering Degree from the École Polytechnique, France, in 2001, the M.Sc. degree in Electrical Engineering from the University of Michigan, Ann Arbor, in 2003, and the Ph.D. degree in Aeronautics and Astronautics from the Massachusetts Institute of Technology, Cambridge, in 2008. He is currently an Associate Professor with the Department of Electrical Engineering, Polytechnique Montreal, Canada. He is a member of GERAD, a multi-university research center on decision analysis. From 2008 to 2012 he was a Postdoctoral Researcher with the GRASP Laboratory at the University of Pennsylvania. In 2018-2019, he was an Alexander von Humboldt Fellow at the Technical University of Munich. His research interests include robust and stochastic control, navigation systems and active perception, networked and multi-agent control systems, with applications to autonomous robots and intelligent infrastructure systems.

Emmanuel Zenou is graduated from “Ecole Normale Supérieure de Paris-Saclay” in 2001 and received a PhD degree from “Ecole Nationale Supérieure de l’Aéronautique et de l’Espace” (SUPAERO) in 2004 in the field of Robotics, Computer Vision and Computer Science. He has a full-time position at ISAE-SUPAERO in the field of Computer Vision and Data Science, in the department of Complex System Engineering (DISC). Since 2017, he is Deputy Director for International Affairs at ISAE-SUPAERO and Deputy Chairman of the Pegasus Network.

Christine Espinosa received the Master in Applied Mathematics in 1982, and the PhD in Theoretical and Computational Mechanics from the University of BORDEAUX I (France, 1991) for the development of a non-linear damage anisotropic behaviour for aeronautical composite materials subjected to low velocity impacts. She worked 9 years in DYNALIS Cie as rapid nonlinear simulations engineer and contributed developing FSI solutions in LS-DYNA®. Full Professor of ISAE-SUPAERO since 2018 in Computational Mechanics for strength analysis of structures under severe multiphysic loads, she received the ‘Palmes Académiques’ in 2019.

Vincent Dubanchet received the M.Sc in aerospace engineering from ISAE-SUPAERO, France, and the M.Sc in electric and electronic engineering from Polytechnique Montréal, Canada, in 2012. In 2016 he obtained the Ph.D. degree from ISAE-SUPAERO and Polytechnique Montréal, for the modeling and control of a flexible space robot to capture tumbling debris. He is currently working as a R&D engineer in Guidance, Navigation & Control and robotics in the R&D department in Thales Alenia Space Cannes. His research interests include in-orbit servicing and autonomous rendezvous.