Consistency of the monocular EKF-SLAM algorithm for 3 different landmark parametrizations

Joan Solà

***Abstract*— We benchmark in this article three different parametrizations for punctual landmarks in monocular 6DOF EKF-SLAM. These parametrizations are *homogeneous points* (HM), *inverse-distance points* (IDP, better known as *inverse-depth*), and the new *anchored homogeneous points* (AHP). The discourse used for describing them is chosen to highlight their differences and similarities, showing that they are just incremental variations of ones with respect to the others. We show for the first time a complete comparison of HP against IDP, two methods that are getting popular, and introduce also for the first time AHP, whose description falls precisely between the other two. The benchmarking is done by running all algorithms on the same data and by using the well-established NEES consistency analysis. Our conclusion is that the new AHP parametrization is the most interesting one for monocular EKF-SLAM (followed by IDP and then HP) because it greatly postpones the apparition of EKF inconsistency.**

I. Introduction

Monocular simultaneous localization and mapping (SLAM) gained popularity back in 2003 thanks to a real-time implementation due to Davison [1]. Davison's technique elegantly solved a great number of problems, but one still remained that occupied researchers on visual SLAM for some years: the problem of landmark initialization. Monocular EKF-SLAM reached maturity with the advent of undelayed initialization techniques, a need firstly stated in 2005 by Solà *et al.* [2], with a preliminary solution based on a previous work in 2004 by Kwok *et al.* [3], and finally solved in 2006 with the inverse-depth landmark parametrization (IDP) due to Montiel *et al.* [4].

The problem of undelayed landmark initialization within monocular EKF-SLAM knows today two main solutions, both of them relying on astute landmark parametrizations: *inverse-distance points* (IDP, better known as inverse-depth, [4], [5]), and *homogeneous points* (HP, [6]). These parametrizations simultaneously fulfill two key objectives: the ability to encode uncertainty up to infinity with one single Gaussian, and the quasi-linearity of the observation functions within all this uncertainty range. These two assets contribute to make undelayed initialization successful with the use of a simple EKF.

Other authors investigated the possibilities of using different estimation techniques. We have seen IDP used in FastSLAM2.0 [7] and UKF [8], [9] frameworks; and methods based on bundle adjustment [10], [11] or on graph theory [12]. These works are often motivated by inconsistency problems and computational burden associated with EKF-SLAM. Without discarding EKF, these two drawbacks can be circumvented with the use of multi-map techniques. The present paper digs deep into the consistency issues.

The aim of this paper is to compare different landmark parametrizations for monocular EKF-SLAM: IDP, HP and a new parametrization, *anchored homogeneous points* (AHP). It also aims at highlighting the similarities between them: we show that AHP is just an anchored version of HP, and that IDP is just a lightened version of AHP. Once these relations are established, the different methods are benchmarked; this allows the reader to correlate the results with the theoretical links presented. The benchmarking is performed with Monte-Carlo simulations, using the well-established normalized estimation error squared (NEES) measure [13] to evaluate consistency. The outcome of this evaluation shows that AHP is the best parametrization in terms of filter consistency, clearly outperforming IDP, and that HP is the worst one. Therefore, an important contribution of this paper is the new AHP parametrization.

Experiments with real images are not presented: IDP and HP have already been demonstrated with real imagery, and our comparison aims at showing performance differences based on the parametrization choice only.

The rest of this article is organized as follows. We detail in Section II the three landmark parametrizations, with their transformation, and perspective projection and back-projection functions. We describe in Section III the initialization and update mechanisms, valid for all parametrizations. In Section IV we give the results of the Monte-Carlo consistency analysis and conclude in Section V with a discussion.

II. Landmark parametrizations

A. Euclidean points (EP)

A Euclidean point $\mathbf{p}$ is trivially coded with three Cartesian coordinates

$$\mathcal{P}_{\mathbb{E}} = \mathbf{p} = \begin{bmatrix} X & Y & Z \end{bmatrix}^\top \in \mathbb{R}^3$$

Transformation to camera frame and pin-hole projection operations resume to

$$\underline{\mathbf{u}} = \mathbf{K}\mathsf{R}^\top(\mathbf{p} - \mathsf{T}) \in \mathbb{P}^2, \tag{1}$$

where $\mathbf{K}$ is the intrinsic matrix, underlined fonts $\underline{\bullet}$ indicate homogeneous coordinates, $\mathsf{R} = \mathsf{R}(\mathsf{Q})$ and T are the rotation matrix and the translation vector defining the camera frame $\mathcal{C} = (\mathsf{T}, \mathsf{Q})$, and Q is a suitable orientation representation (we use quaternions).

J. Solà is at University of Toulouse, LAAS-CNRS, France. jsola@laas.fr

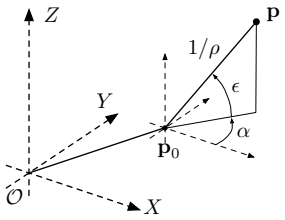


Fig. 1. Inverse-distance point (IDP) parametrization. The anchor point $\mathbf{p}_0$ corresponds to the optical center at initialization time. We encode the unmeasured distance with its inverse ρ.

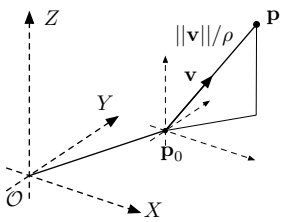


Fig. 2. Anchored homogeneous point (AHP) parametrization. We define the ray's direction with a vector $\mathbf{v}$ that, together with the inverse distance ρ, constitutes a homogeneous point referenced at the anchor point $\mathbf{p}_0$. In fact, there is no need for $\mathbf{v}$ to be unitary, in which case ρ is not inverse-distance but proportional to inverse-distance: $d = ||\mathbf{p} - \mathbf{p}_0|| = ||\mathbf{v}||/\rho$.

Euclidean points lead to severely non-linear observation functions in bearings-only systems and are not suited for undelayed initialization, as it has been extensively demonstrated [2], [3], [6], [7], [14] and most particularly [5]. The parametrizations that follow mitigate this problem and can be used for undelayed initialization with just a few precautions.

B. Inverse-distance points (IDP)

An "inverse-distance" point[1] (IDP, Fig. 1) [5] is coded by a 6-vector containing the Euclidean optical center at initialization time, $\mathbf{p}_0 = (x_0, y_0, z_0)$, elevation and azimuth angles defining the direction of the initial optical ray, (ε, α), and the inverse of the Euclidean distance from $\mathbf{p}_0$ to the 3D point $\mathbf{p}$, denoted by ρ:

$$\begin{aligned}\mathcal{P}_{\mathbb{ID}} &= \begin{bmatrix}\mathbf{p}_0^\top & \varepsilon & \alpha & \rho\end{bmatrix}^\top \\ &= \begin{bmatrix}x_0 & y_0 & z_0 & \varepsilon & \alpha & \rho\end{bmatrix}^\top \in \mathbb{R}^6\end{aligned} \tag{2}$$

We will refer to the initial optical center $\mathbf{p}_0$ as the *anchor point* of the landmark. An IDP refers to the following EP:

$$\mathbf{p} = \mathbf{p}_0 + \mathbf{v}^*(\varepsilon, \alpha)/\rho \tag{3}$$

where $\mathbf{v}^*(\varepsilon, \alpha)$ is a unit vector in the direction of (ε, α),

$$\mathbf{v}^*(\varepsilon, \alpha) = \begin{bmatrix}\cos(\varepsilon)\cos(\alpha) & \cos(\varepsilon)\sin(\alpha) & \sin(\varepsilon)\end{bmatrix}^\top. \tag{4}$$

Transformation to camera frame and pin-hole projection operations resume to

$$\underline{\mathbf{u}} = \mathbf{K}\mathsf{R}^\top\big(\mathbf{v}^*(\varepsilon, \alpha) - \rho(\mathsf{T} - \mathbf{p}_0)\big). \tag{5}$$

The back-projection and transformation composition necessary for initialization is performed with

$$\mathcal{P}_{\mathbb{ID}} = \begin{bmatrix}\mathbf{p}_0 \\ (\varepsilon, \alpha) \\ \rho\end{bmatrix} = \begin{bmatrix}\mathsf{T} \\ \upsilon^*(\mathsf{R}\mathbf{K}^{-1}\underline{\mathbf{u}}) \\ \rho^{\mathcal{C}}\end{bmatrix}, \tag{6}$$

where $\upsilon^*(\mathbf{v})$ gives elevation and azimuth angles (ε, α) of a director vector $\mathbf{v} = (u, v, w)$,

$$\begin{bmatrix}\varepsilon \\ \alpha\end{bmatrix} = \upsilon^*(u, v, w) = \begin{bmatrix}\arctan(w/\sqrt{u^2 + v^2}) \\ \arctan(v/u)\end{bmatrix}. \tag{7}$$

The inverse-distance parameter $\rho^{\mathcal{C}}$ is defined in the camera frame at initialization time. It must be provided as prior.

[1] In this article we will refer to the originally named "inverse depth" points as *inverse-distance* points, and will use the invariant abbreviation IDP.

C. Anchored homogeneous points (AHP)

IDP points can be parametrized somewhat differently by encoding the optical ray's direction with a vector $\mathbf{v} = (u, v, w)$, avoiding the need for the non-linear transformations (4) and (7). When this vector is unitary, appending the inverse of the distance ρ to it results in a homogeneous point $(u, v, w, \rho) \in \mathbb{P}^3$. This leads to the anchored homogeneous point (AHP, Fig. 2), parametrized with the 7-vector

$$\begin{aligned}\mathcal{P}_{\mathbb{AH}} &= \begin{bmatrix}\mathbf{p}_0^\top & \mathbf{v}^\top & \rho\end{bmatrix}^\top \\ &= \begin{bmatrix}x_0 & y_0 & z_0 & u & v & w & \rho\end{bmatrix}^\top \in \mathbb{R}^7\end{aligned} \tag{8}$$

It is worth noticing that a homogeneous point $(\mathbf{v}, \rho)$ does not require $\mathbf{v}$ to be a unit vector. If it is not, the parametrization is absolutely valid but ρ is then not the inverse distance $1/d$ but something proportional to it, *i.e.*, $\rho = \|\mathbf{v}\|/d$.

An AHP refers to the following EP:

$$\mathbf{p} = \mathbf{p}_0 + \mathbf{v}/\rho. \tag{9}$$

Transformation to camera frame and projection resume to

$$\underline{\mathbf{u}} = \mathbf{K}\mathsf{R}^\top\big(\mathbf{v} - \rho(\mathsf{T} - \mathbf{p}_0)\big) \in \mathbb{P}^2. \tag{10}$$

The back-projection and transformation composition is done with

$$\mathcal{P}_{\mathbb{AH}} = \begin{bmatrix}\mathbf{p}_0 \\ \mathbf{v} \\ \rho\end{bmatrix} = \begin{bmatrix}\mathsf{T} \\ \mathsf{R}\mathbf{K}^{-1}\underline{\mathbf{u}} \\ \rho^{\mathcal{C}}\end{bmatrix}, \tag{11}$$

where $\rho^{\mathcal{C}}$ must be provided as prior; its relation to distance d is given by $\rho^{\mathcal{C}} = \|\mathbf{K}^{-1}\underline{\mathbf{u}}\|/d$.

D. Homogeneous points (HP)

Homogeneous points have the interesting property of presenting a bi-linear transformation equation:

$$\underline{\mathbf{p}} = \mathbf{H}\underline{\mathbf{p}}^{\mathcal{C}} \triangleq \begin{bmatrix}\mathsf{R}(\mathsf{Q}) & \mathsf{T} \\ 0 & 1\end{bmatrix}\underline{\mathbf{p}}^{\mathcal{C}}. \tag{12}$$

When the uncertainties on the camera position $\mathcal{C} = (\mathsf{T}, \mathsf{Q})$ are small, we can consider the motion matrix $\mathbf{H}$ to be deterministic, and therefore (12) to be exactly linear. In this case Gaussian uncertainties are transformed exactly, so it does not matter where our landmarks are anchored at. We are free to re-anchor the AHP (8) at the origin with

$$\begin{bmatrix}\mathbf{v} \\ \rho\end{bmatrix} \leftarrow \begin{bmatrix}\mathbf{I}_{3\times3} & \mathbf{p}_0 \\ \mathbf{0}_{\mathbf{1}\times\mathbf{3}} & 1\end{bmatrix}\begin{bmatrix}\mathbf{v} \\ \rho\end{bmatrix}, \tag{13}$$

TABLE I

SUMMARY OF LANDMARK PARAMETRIZATIONS AND THEIR MAIN MANIPULATIONS

Lmk	parameters	size	transformation	projection	transformation + projection $h()$	back-projection + transf. $g()$
EP	$\mathcal{P}_{\mathbb{E}} = \mathbf{p}$	3	$\mathbf{p} = \mathsf{R}\mathbf{p}^{\mathcal{C}} + \mathsf{T}$	$\underline{\mathbf{u}} = \mathbf{K}\mathbf{p}^{\mathcal{C}}$	$\underline{\mathbf{u}} = \mathbf{K}\mathsf{R}^\top(\mathbf{p} - \mathsf{T})$	$\mathbf{p} = t\mathsf{R}\mathbf{K}^{-1}\underline{\mathbf{u}} + \mathsf{T}$
HP	$\mathcal{P}_{\mathbb{H}} = \underline{\mathbf{p}} = (\mathbf{v}, \rho)$	4	$\underline{\mathbf{p}} = \mathbf{H}\underline{\mathbf{p}}^{\mathcal{C}}$	$\underline{\mathbf{u}} = \mathbf{K}\mathbf{v}^{\mathcal{C}}$	$\underline{\mathbf{u}} = \mathbf{K}\mathsf{R}^\top(\mathbf{v} - \rho\mathsf{T})$	$\mathcal{P}_{\mathbb{H}} = \underline{\mathbf{p}} = \mathbf{H}\begin{bmatrix}\mathbf{K}^{-1}\underline{\mathbf{u}} \\ \rho^{\mathcal{C}}\end{bmatrix}$
AHP	$\mathcal{P}_{\mathbb{AH}} = (\mathbf{p}_0, \mathbf{v}, \rho)$	7			$\underline{\mathbf{u}} = \mathbf{K}\mathsf{R}^\top\big(\mathbf{v} - \rho(\mathsf{T} - \mathbf{p}_0)\big)$	$\mathcal{P}_{\mathbb{AH}} = \begin{bmatrix}\mathsf{T} \\ \mathsf{R}\mathbf{K}^{-1}\underline{\mathbf{u}} \\ \rho^{\mathcal{C}}\end{bmatrix}$
IDP	$\mathcal{P}_{\mathbb{ID}} = (\mathbf{p}_0, \varepsilon, \alpha, \rho)$	6			$\underline{\mathbf{u}} = \mathbf{K}\mathsf{R}^\top\big(\mathbf{v}^* - \rho(\mathsf{T} - \mathbf{p}_0)\big)$	$\mathcal{P}_{\mathbb{ID}} = \begin{bmatrix}\mathsf{T} \\ \upsilon^*(\mathsf{R}\mathbf{K}^{-1}\underline{\mathbf{u}}) \\ \rho^{\mathcal{C}}\end{bmatrix}$

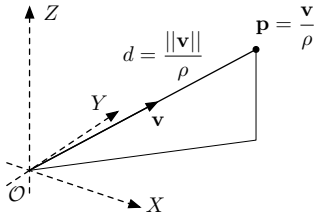


Fig. 3. Homogeneous point (HP) parametrization. The homogeneous part of AHP is anchored at the origin, and the anchor suppressed. There is no need for $\mathbf{v}$ to be a unit vector.

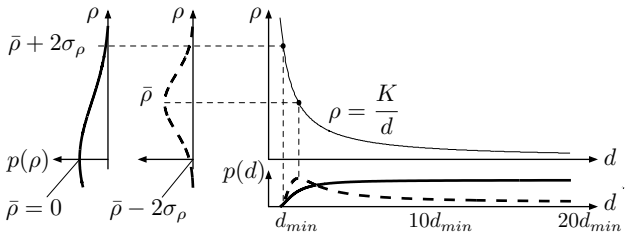


Fig. 4. Inverse-distance PDF. A Gaussian $p(\rho) = \mathcal{N}(\rho - \bar{\rho}, \sigma_\rho^2)$ is defined in inverse-distance (vertical axes). We have ample choice: in one extreme (dashed) we may define it so that $\bar{\rho} - 2\sigma_\rho = 0$; the other extreme (solid) takes $\bar{\rho} = 0$. In all cases, we have $(\bar{\rho} + 2\sigma_\rho) = K/d_{min}$. They result in PDFs in distance (bottom) that cover from a minimal distance d_{min} to infinity. K is just a proportionality constant, *e.g.* $K = 1$ for IDP, and $K = \|\mathbf{K}^{-1}\underline{\mathbf{u}}\|$ for AHP and HP. We can also normalize $\mathbf{K}^{-1}\underline{\mathbf{u}}$ at initialization time and take $K = 1$, in which case ρ is exactly equal to inverse-distance.

which allows us to remove the anchor $\mathbf{p}_0$. This leads to a purely homogeneous point (HP, Fig. 3), which has already been studied in [6]:

$$\mathcal{P}_{\mathbb{H}} = \begin{bmatrix}\mathbf{v}^\top & \rho\end{bmatrix}^\top = \begin{bmatrix}u & v & w & \rho\end{bmatrix}^\top \in \mathbb{R}^4. \tag{14}$$

A HP refers to the following EP:

$$\mathbf{p} = \mathbf{v}/\rho. \tag{15}$$

Transformation to camera frame and projection resume to

$$\underline{\mathbf{u}} = \mathbf{K}\mathsf{R}^\top(\mathbf{v} - \rho\mathsf{T}) \in \mathbb{P}^2. \tag{16}$$

The back-projection and transformation composition is done with

$$\mathcal{P}_{\mathbb{H}} = \underline{\mathbf{p}} = \begin{bmatrix}\mathbf{v} \\ \rho\end{bmatrix} = \mathbf{H}\begin{bmatrix}\mathbf{K}^{-1}\underline{\mathbf{u}} \\ \rho^{\mathcal{C}}\end{bmatrix}, \tag{17}$$

where $\rho^{\mathcal{C}}$ must be provided as prior; its relation to initial distance $d^{\mathcal{C}}$ is given by $\rho^{\mathcal{C}} = \|\mathbf{K}^{-1}\underline{\mathbf{u}}\|/d^{\mathcal{C}}$. Once transformed to the global frame with $\mathbf{H}$, this meaning of $\rho^{\mathcal{C}}$ is lost and therefore not valid for ρ.

E. Final comment

We have presented three parametrizations and shown the links between them. We have chosen to start by IDP purely because of chronological reasons. The reader should be able to construct a discourse in the inverse order: start by HP, well known for their interesting properties in vision, then AHP as an anchored version of HP, then IDP as a lightened version of AHP. We summarize in Table I all parametrizations with their main manipulation expressions. This should help building a coherent picture of the parametrizations suited for monocular EKF-SLAM.

III. INITIALIZATION AND UPDATES

A. Initialization

Undelayed landmark initialization in EKF-SLAM with partial measurements (such as monocular measurements) mimics the algorithm for full measurements by incorporating the non-measured magnitudes as priors:

1) Identify the known magnitudes: measurement $\mathbf{u} \sim \mathcal{N}\{\mathbf{y}, \mathbf{R}\}$ and map $X \sim \mathcal{N}\{\bar{X}, \mathbf{P}\}$, where
$$X = \begin{bmatrix}\mathcal{C} \\ \mathcal{M}\end{bmatrix}, \quad \bar{X} = \begin{bmatrix}\bar{\mathcal{C}} \\ \bar{\mathcal{M}}\end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix}\mathbf{P}_{\mathcal{C}\mathcal{C}} & \mathbf{P}_{\mathcal{C}\mathcal{M}} \\ \mathbf{P}_{\mathcal{M}\mathcal{C}} & \mathbf{P}_{\mathcal{M}\mathcal{M}}\end{bmatrix},$$
with $\mathcal{C} = (\mathsf{T}, \mathsf{Q})$ the camera frame and $\mathcal{M}$ the set of mapped landmarks.
2) Define a Gaussian prior for the non-measured inverse distance, $\rho^{\mathcal{C}} \sim \mathcal{N}\{\bar{\rho}^{\mathcal{C}}, \sigma_{\rho^{\mathcal{C}}}^2\}$, see Fig. 4.
3) Back-project the Gaussian measurement; get landmark mean and Jacobians
$$\bar{\mathcal{P}} = g(\bar{\mathcal{C}}, \mathbf{y}, \bar{\rho}^{\mathcal{C}})$$
$$\mathbf{G}_{\mathcal{C}} = \left.\frac{dg}{d\mathcal{C}}\right|_{\bar{\mathcal{C}}, \mathbf{y}, \bar{\rho}^{\mathcal{C}}}, \mathbf{G}_{\mathbf{u}} = \left.\frac{dg}{d\mathbf{u}}\right|_{\bar{\mathcal{C}}, \mathbf{y}, \bar{\rho}^{\mathcal{C}}}, \mathbf{G}_{\rho} = \left.\frac{dg}{d\rho}\right|_{\bar{\mathcal{C}}, \mathbf{y}, \bar{\rho}^{\mathcal{C}}}$$
with $g(\mathcal{C}, \mathbf{u}, \rho^{\mathcal{C}})$ one of the back-projection functions in Table I, $\mathcal{C} = (\mathsf{T}, \mathsf{Q})$, $\mathsf{R} = \mathsf{R}(\mathsf{Q})$ and $\underline{\mathbf{u}} = [\mathbf{u}^\top \; 1]^\top$.
4) Compute landmark co- and cross-variances
$$\mathbf{P}_{\mathcal{P}\mathcal{P}} = \mathbf{G}_{\mathcal{C}}\mathbf{P}_{\mathcal{C}\mathcal{C}}\mathbf{G}_{\mathcal{C}}^\top + \mathbf{G}_{\mathbf{u}}\mathbf{R}\mathbf{G}_{\mathbf{u}}^\top + \mathbf{G}_{\rho}\sigma_{\rho^{\mathcal{C}}}^2\mathbf{G}_{\rho}^\top$$
$$\mathbf{P}_{\mathcal{P}X} = \mathbf{G}_{\mathcal{C}}\mathbf{P}_{\mathcal{C}X}$$

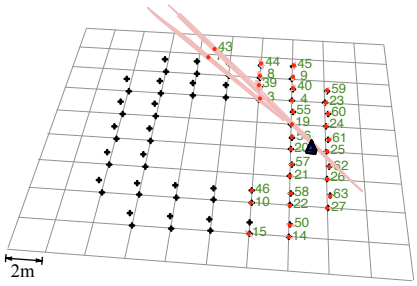


Fig. 5. Simulated 3D environment for 6-DOF monocular SLAM.

with $\mathbf{P}_{\mathcal{C}X} = [\mathbf{P}_{\mathcal{C}\mathcal{C}}\ \mathbf{P}_{\mathcal{C}\mathcal{M}}]$.

5) Augment the SLAM map

$$\bar{X} \leftarrow \begin{bmatrix} \bar{X} \\ \bar{\mathcal{P}} \end{bmatrix}, \qquad \mathbf{P} \leftarrow \begin{bmatrix} \mathbf{P} & \mathbf{P}_{\mathcal{P}X}^\top \\ \mathbf{P}_{\mathcal{P}X} & \mathbf{P}_{\mathcal{P}\mathcal{P}} \end{bmatrix}.$$

B. Updates

Updates follow the standard EKF-SLAM formulation. The observation functions $\mathbf{u} = h(\mathcal{C}, \mathcal{P})$ are the composition of the ones in Table I with the homogeneous-to-Euclidean transform: if $\underline{\mathbf{u}} = [u, v, w]^\top$ then $\mathbf{u} = [u/w, v/w]^\top$.

IV. Consistency analysis

We benchmark HP, AHP and IDP for filter consistency using the same simulated scenario, the same software and the same seeds for the random generator. A description of the simulation conditions and the benchmarking methods follows, and results are given at the end of this section.

A. Simulated scenario

We simulate a robot performing a circular trajectory in an area of $12m \times 12m$, populated with 72 landmarks forming a cloister (Fig. 5). The robot receives noisy control inputs which are used for the prediction stage of the EKF, fixing the scale factor. One noisy image per control step is gathered with a single camera heading forward. Two sets of parameters have been used for the tests (the nominal and perturbation levels of all these magnitudes, together with the inverse-distance priors used, are all summarized in Table II). In the first set, the robot makes two turns to the cloister (800 frames are processed). The second set uses smaller odometry increments and perturbations, and the trajectory is limited to one quarter of a turn (200 frames).

B. Software and SLAM algorithm

We have made available the software used for simulations [15]. It consists in a 6-DOF EKF-SLAM system written in MATLAB®, with simulation and 3D graphics capabilities.

The algorithm is organized as an active-search-based SLAM [16], which allows us to optimize information gain with a limited number of updates per frame. At each frame, we perform updates to the 10 most informative landmarks. We also attempt to initialize one landmark per frame. Inconsistent and unstable landmarks are deleted from the map to avoid map corruption.

TABLE II

Simulation parameters for all experiments

Concept	Param.	Set 1	Set 2
Pose step	$(\Delta X, \Delta\psi)$	$(8\text{cm}, 0.9°)$	$(4\text{cm}, 0.45°)$
Lin. noise	$(\sigma_X, \sigma_Y, \sigma_Z)$	1cm	0.5cm
Ang. noise	$(\sigma_\phi, \sigma_\theta, \sigma_\psi)$	0.1°	0.05°
Img. size		640×480 pix	
Focal	(α_u, α_v)	320 pix, $FOV = 90°$	
Pix. noise	$\sigma_{\mathbf{u}}$	1 pix	
$\rho^{\mathcal{C}}$ prior	$(\bar{\rho}^{\mathcal{C}}, \sigma_{\rho^{\mathcal{C}}})$	$(0.01, 0.5)$ m^{-1}	id. + $(1, 1)$ m^{-1}

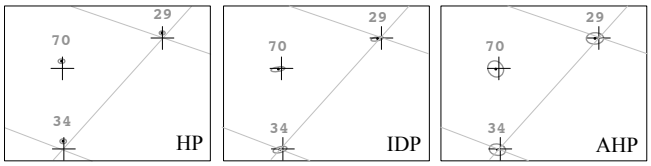


Fig. 6. 3D view of some landmark 3σ estimates at the end of the first loop. Inconsistency comes from covariance overestimation rather than mean errors. See the accompanying video.

C. Normalized estimation error squared (NEES)

Here we follow [13]. We analyze filter consistency using the normalized estimation error squared (NEES). When ground truth about a variable $\mathbf{x}_k$ is known, the NEES of its estimate $\mathcal{N}\{\hat{\mathbf{x}}_k, \mathbf{P}_k\}$ is defined at each time k by

$$\epsilon_k = (\mathbf{x}_k - \hat{\mathbf{x}}_k)^\top \mathbf{P}_k^{-1} (\mathbf{x}_k - \hat{\mathbf{x}}_k). \tag{18}$$

Under the hypothesis of consistent filtering of a linear-Gaussian system, ϵ_k obeys a χ^2 distribution with $\dim(\mathbf{x}_k)$ degrees of freedom (DOF), noted $\chi^2_{\dim(\mathbf{x})}$, whose expectation over an increasing number of runs converges to the state dimension, $\mathbb{E}[\epsilon_k] = \dim(\mathbf{x}_k)$. The linear-Gaussian hypothesis can then be statistically evaluated by means of a χ^2 acceptance test over a set of $N < \infty$ Monte-Carlo runs.

Given N Monte-Carlo runs, $\sum_{i=1}^N \epsilon_{ik}$ obeys a $\chi^2_{N\dim(\mathbf{x})}$ distribution. The bounds of the *double-sided* 95% *probability concentration region* are given by the $\chi^2_{N\dim(\mathbf{x})}$ values corresponding to tail probabilities of 2.5% and 97.5%. For 6-DOF SLAM and $N = 25$ runs, we have the lower and upper bounds $\{\underline{\nu}; \overline{\nu}\} = \chi^2_{(6\times25)}(1 - 0.025; 1 - 0.975) = \{117.985; 185.800\}$.

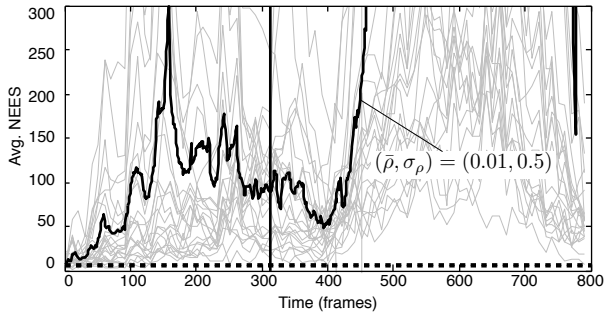
The average NEES is computed as

$$\bar{\epsilon}_k \triangleq \frac{1}{N} \sum_{i=1}^N \epsilon_{ik}. \tag{19}$$

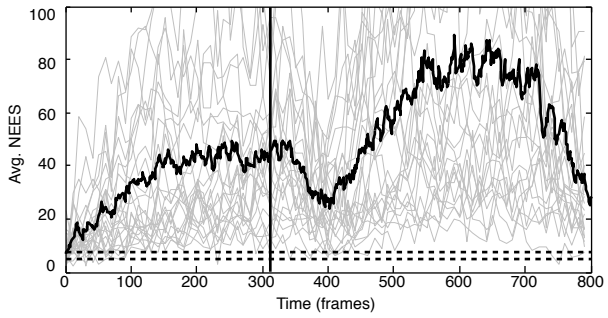
We compare the average NEES against $\{\underline{\nu}/N; \overline{\nu}/N\} = \{4.719; 7.432\}$. If the average NEES is below the lower bound for some significant amount of time, the filter is conservative. If it is above the upper bound, the filter is optimistic and therefore inconsistent.

D. Results

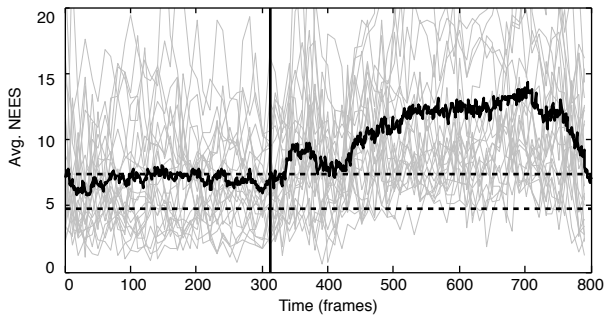
We provide an accompanying video showing the three methods running in parallel. The differences in behavior are



(a) HP

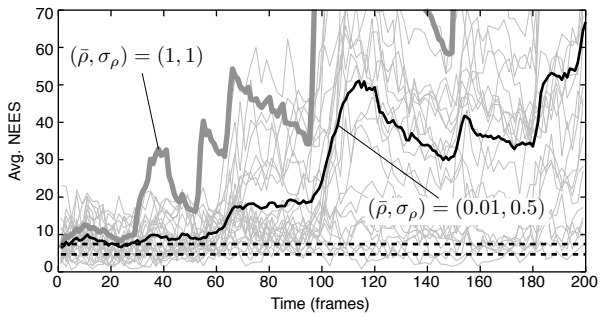


(b) IDP

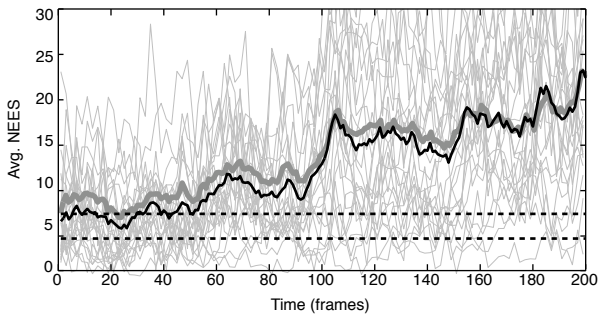


(c) AHP

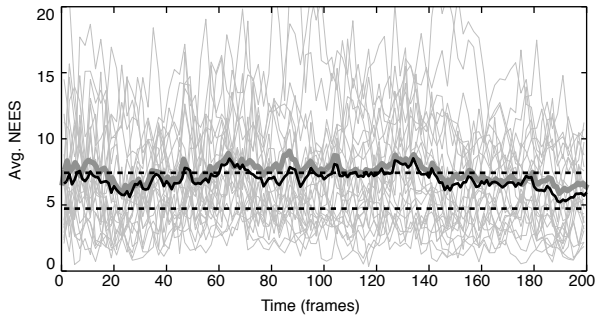
Fig. 7. Average NEES of the 6-DOF vehicle pose $[x, y, z, \phi, \theta, \psi]^\top$ over 25 runs for 800 frames (2 turns) and parameters of Set 1. The thin gray lines are the 25 individual NEES. The thick line is the average NEES. Dashed horizontal lines at abscissas 4.719 and 7.432 delimit the 95% consistency region. The vertical line marks the loop closure at frame 308.



(a) HP



(b) IDP



(c) AHP

Fig. 8. Average NEES over 25 runs for 200 frames and alternative set of conditions (1/4 turn) with parameters of Set 2 and 10 initializations in the first frame. *Thick gray:* initialization with an alternative prior $(\bar{\rho}, \sigma_\rho) = (1, 1)$.

not easily visible in the 3D movies, and we need to zoom in to appreciate incorrect operation (Fig. 6: IDP and HP estimate too small covariances). However, their NEES behavior is radically different (Fig. 7, please note the different vertical scales):

HP behaves poorly. Of the 25 runs, one diverged, and 35 landmarks had to be deleted due to inconsistent observations (22 of which during the divergent run).

IDP shows better performance but also escapes consistency very quickly. No run diverged but inconsistent observations triggered landmark deletion in two occasions.

AHP behaves consistently, certainly with a slight tendency to inconsistency, until shortly after the first loop closure. During the second turn the filter is inconsistent but it does not seem to degrade too quickly. No landmarks were declared inconsistent.

We tuned the algorithms with the second set of parameters in order to improve linearity: odometry steps and noise are cut in half, and the filter is bootstrapped with 10 landmarks being initialized at the first frame. Here, we focus on the first quarter of the loop to see the moment when the filters loose consistency. The results in Fig. 8 show no significant improvement with respect to those of Set 1 (these 200 frames correspond to the first 100 frames in Fig. 7): HP is just not good, IDP starts fine but only keeps track until frame 50, and AHP is again the only one to behave consistently.

A third test consisted in selecting a different prior for the unmeasurable inverse-distance. The gray superimposed plots in Fig. 8 show that IDP and AHP are not very sensitive to large variations of these parameters, while the contrary must be said for HP. It seems, even if for AHP and IDP

the difference is small, that the filter behaves better with landmarks initialized at (or close to) infinity ($\bar{\rho}^{\mathcal{C}} = 0.01\mathrm{m}^{-1}$) than at some close distance ($\bar{\rho}^{\mathcal{C}} = 1\mathrm{m}^{-1}$).

V. Discussion and conclusion

We are dealing with 3 parametrizations. One is not anchored (HP); the other two are anchored (IDP and AHP). Thanks to the cross-correlations stored in the EKF, the anchor allows the filter to account for accumulated errors only from the anchor to the current position (terms $(\mathsf{T} - \mathbf{p}_0)$ in Table I), not from the origin of coordinates (term T). This is consistent with HP performing clearly worse than IDP and AHP.

We expected AHP to be better than IDP, but we could not tell to which extent. Our experiments showed that the improvement is very significant. We see two reasons for this:

First, the ray direction in AHP is a 3D vector $\mathbf{v}$ that, importantly, is not forced to unity. The fact that its norm can evolve during filtering allows the filter to work more relaxed (there is redundancy or lack of constraints). So when an update occurs, the correction effect can be shared between several dimensions, including of course ρ. In IDP there is no such redundancy and the filter works more "constrained". When it comes to deal with non-linearity, these constraints contribute to larger errors. It is therefore important not to normalize the vector $\mathbf{v}$ at each frame. We might, though, normalize it at initialization time to render ρ more meaningful as true inverse-distance, thus making the initialization algorithm more readable.

Second, the transformation equations in AHP are more linear than those in IDP, because in IDP we have the trigonometric functions (4) and (7). In fact, AHP only differs from IDP in two single lines of code: those performing such equations. This point is, we feel, not very important, because these angles are well observed from the first observation and therefore the trigonometric functions can be considered quite linear inside the uncertainty region, which is small. However, its effect can only contribute to increase linearization errors.

The differences in consistency between IDP and AHP are indeed surprising. IDP has been extensively used with success already for three years, and one could imagine other reasons for our relatively poor IDP results. For example, one could conjecture that our implementation using quaternions for orientation could be at the base of our IDP inconsistency (this is a hypothesis drawn only from the fact that we did not observe this behavior with another implementation using Euler angles – but we were not using the NEES tool either). Should this be the case, we would have found the same inconsistency in AHP, and we did not. Our conclusion is again that AHP positively improves on IDP performance.

Regarding the increase in computational costs derived from a larger parametrization (size 7 instead of 6), we point here that the linearity measure for IDP [5] can be applied to AHP as-is to trigger the landmark reparametrization. Reparametrization to Euclidean has virtually no effect on NEES results (something we did not show here for space reasons) – though a very small degradation can be observed. This degradation is of the order of the one observed when changing the prior values of $\rho^{\mathcal{C}}$, which we saw in Fig. 8. The gain in computational power, however, largely compensates for it.

Regarding terminology, IDP could be renamed AMPP (*Anchored Modified Polar Point*): it is anchored, and it is in polar coordinates except for the radius which is inverse-radius (therefore the "modified polar"). This would produce a consistent picture, where the concept of "inverse-distance", as we have seen, is shared among all the parametrizations. To complete the picture, we notice that the AMPP's un-anchored counterpart, *Modified Polar Point* (MPP), not presented in this paper, had already been treated in the 80's in the bearing-only tracking literature with similar problematic and justification [17]. Its use in monocular EKF-SLAM is not recommended: it presents a singularity at the origin, and if we draw the correct conclusions from the present paper, it should behave even worse than HP, as it happens to AMPP (*i.e.* IDP) with respect to AHP.

References

[1] A. J. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *Int. Conf. on Computer Vision*, vol. 2, Nice, October 2003, pp. 1403–1410.

[2] J. Solà, A. Monin, M. Devy, and T. Lemaire, "Undelayed initialization in bearing only SLAM," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Edmonton, Canada, 2005, pp. 2499–2504.

[3] N. M. Kwok and G. Dissanayake, "An efficient multiple hypothesis filter for bearing-only SLAM," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Sendai, Japan, 2004.

[4] J. Montiel, J. Civera, and A. J. Davison, "Unified inverse depth parametrization for monocular SLAM," in *Robotics: Science and Systems*, Philadelphia, USA, August 2006.

[5] J. Civera, A. Davison, and J. Montiel, "Inverse depth parametrization for monocular SLAM," *IEEE Trans. on Robotics*, vol. 24, no. 5, 2008.

[6] D. Marzorati, M. Matteucci, D. Migliore, and D. G. Sorrenti, "Monocular SLAM with inverse scaling parametrization," in *Proc. of the British Machine Vision Conference*, Leeds, 2008.

[7] E. Eade and T. Drummond, "Scalable monocular SLAM," *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 469–476, 2006.

[8] N. Sunderhauf, S. Lange, and P. Protzel, "Using the unscented kalman filter in mono-SLAM with inverse depth parametrization for autonomous airship control," in *IEEE Int. Workshop on Safety, Security and Rescue Robotics*, Rome, 2007.

[9] S. A. Holmes, G. Klein, and D. W. Murray, "A square root UKF for visual monoSLAM," in *IEEE Int. Conf. on Robotics and Automation*, Pasadena, 2008.

[10] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *IEEE and ACM Int. Symp. on Mixed and Augmented Reality*, 2007.

[11] K. Konolige and M. Agrawal, "FrameSLAM: From bundle adjustment to real-time visual mapping," *IEEE Trans. on Robotics*, vol. 24, no. 5, pp. 1066–1077, Oct. 2008.

[12] E. Eade and T. Drummond, "Monocular slam as a graph of coalesced observations," in *IEEE Int. Conf. on Computer Vision*, 2007.

[13] T. Bailey, J. Nieto, J. Guivant, M. Stevens, and E. Nebot, "Consistency of the EKF-SLAM algorithm," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Beijing, China, October 2006, pp. 3562–3568.

[14] T. Bailey, "Constrained initialisation for bearing-only SLAM," in *Int. Conf. on Robotics and Automation*, vol. 2, 2003, pp. 1966–1971.

[15] J. Solà, D. Marquez, and J. M. Codol. (2009) An EKF-SLAM toolbox for MATLAB. [Online]. Available: http://www.laas.fr/~jsola/Joan%20Sola/eng/Joan%20Sola.html#toolbox

[16] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1052–1067, June 2007.

[17] V. Aidala and S. Hammel, "Utilization of modified polar coordinates for bearings-only tracking," *IEEE Transactions on Automatic Control*, vol. 28, no. 3, pp. 283 – 294, March 1983.