2009 IEEE International Conference on Robotics and Automation
Kobe International Conference Center
Kobe, Japan, May 12-17, 2009

# On the Use of Inverse Scaling in Monocular SLAM

Daniele Marzorati, Matteo Matteucci, Davide Migliore, Domenico G. Sorrenti

*Abstract*— **Recent works have shown that it is possible to solve the Simultaneous Localization And Mapping problem using an Extended Kalman Filter and a single perspective camera. The principal drawback of these works is an inaccurate modeling of measurement uncertainties, which therefore causes inconsistencies in the filter estimations. A possible solution to proper uncertainty modeling is the Unified Inverse Depth parametrization. In this paper we propose the Inverse Scaling parametrization that still allows an un-delayed initialization of features, while reducing the number of needed parameters and simplifying the measurement model. This novel approach allows a better uncertainty modeling of both low and high parallax features and reduces the likelihood of inconsistencies. Experiments in simulation demonstrate that the use of the Inverse Scaling solution improves the performance of the monocular EKF SLAM filter when compared with the Unified Inverse Depth approach; experiment on real data confirm the applicability of the idea.**

## I. INTRODUCTION

Monocular cameras are in widespread use for SLAM, as they are simple and low power sensors that allows to estimate the bearing of interest points and, by means of camera motion and triangulation, the whole 3D structure of the environment [1]. A relevant issue in this case is the initialization and the uncertainty modeling of the 3D elements in the map: from a single frame we cannot estimate the depth of the features, and measurements are affected by uncertainties that strongly depend on the observer-to-feature distance.

In their work, Davison et al. [1], using an extended Kalman filter (EKF) to perform a real-time 6 DoF SLAM, used a non parametric approach to initialize the feature depth and bounded the maximum feature depth to about 5m. The depth of an observed feature is first estimated using a particle filter and the feature, once its distribution is close to normal, is used in a EKF-based SLAM filter. Unfortunately, this delayed use can cause a loss of information; in fact having a landmark in the map, even without knowing its distance, allows immediate use of pure bearing information. To avoid this delay and to exploit low-parallax features, Solà et al.[2] proposed to maintain several depth hypotheses combined in a Gaussian Sum Filter, to cover the distribution along the whole ray to the feature. An alternative solution for both un-delayed initialization and depth uncertainty modeling was introduced in [3] and [4]. They showed that the use of inverse depth parameterizations make the observation model nearly linear (at least for small camera displacements), while reducing both non-Gaussian-ness of depth measurement and EKF linearization. In this way, it is possible to model the uncertainty as Gaussian and use EKF filtering, without delay. In [5] it was suggested to base on recursive estimation in local coordinate frames and an iterative graph optimization, obtaining a nearly linear observation model. At the same time Clemente et al. [6] demonstrated that a different solution to filter inconsistencies is to use a Hierarchical map approach that, combined with the Joint Compatibility test, allows to perform a mapping of a large loop.

Starting from these results, our aim is a novel parametrization that allows not only to estimate the depth of the features without delay, but also to obtain a good accuracy in uncertainty modeling for both low and high parallax features, therefore improving the stability of the monocular SLAM filter. In the next section we introduce the Inverse Scaling parametrization, studying the linearity of the measurement model in comparison with the Unified Inverse Depth (UID hereafter) [3] approach. A complete EKF SLAM algorithm using inverse scaling parametrization is presented in the following sections. In section IV we validate our proposal on simulated and real data, comparing the results with the solution presented in [3].

## II. THE INVERSE SCALING PARAMETRIZATION

As proposed in [3] and [4], it is possible to improve the performance of a monocular EKF SLAM adopting an inverse depth parametrization and thus allowing not only an un-delayed initialization of features, but also a non-linearity reduction of the observation model. The latter result can be confirmed by analyzing the linearity of a simplified measurement equation, as showed in [7], and it can be generalized for scene points that are not on the principal axis of the camera. Let us consider Figure 1, which sketches two cameras, with the same focal length, observing a generic point in the scene. If we consider that this point is not laying on the two cameras principal axis (i.e., the general case), the angles $\theta_0$ and $\theta_1$ will not be zero. To include this information in the equation of [7] we can change the equation representing the location error for the first camera as follows:

$$D = \frac{1}{\rho_0 - \rho} cos(\theta_0), \quad d_0 = \frac{1}{\rho_0} cos(\theta_0), \tag{1}$$

$$d = D - d_0 = \frac{\rho}{\rho_0(\rho_0 - \rho)} cos(\theta_0), \tag{2}$$

where $\rho_0$ is the UID of the feature, $\rho \sim N(0, \sigma_\rho^2)$ is the Gaussian depth uncertainty, $d$ is the point's location error projected on the $y$ axis w.r.t the first camera, $d_0$ is the real

D. Marzorati and D. G. Sorrenti are with Università di Milano - Bicocca, Building U14, v.le Sarca 336, 20126, Milano, Italy{marzorati, sorrenti}@disco.unimib.it

M. Matteucci and D. Migliore are with Politecnico di Milano, via Ponzio 34/5, 20133, Milano, Italy{matteucci, migliore}@elet.polimi.it

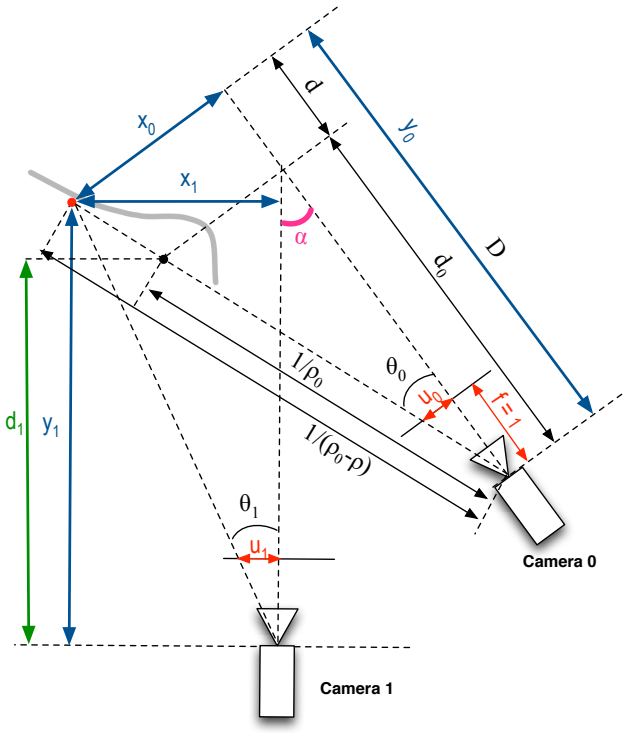



Fig. 1. Modeling the uncertainty propagation from a generic scene point to the image camera. In gray we represents the skewed uncertainty distribution of the measurement.

feature position projected on the $y$ axis and $D$ is position with uncertainty. Considering the second camera pose, we can estimate the image of the scene point:

$$x_1 = \frac{\rho_0 \sin(\theta_0)\cos(\alpha)+\rho\ \cos(\theta_0)\sin(\alpha)}{\rho_0(\rho_0-\rho)}, \quad (3)$$

$$y_1 = d_1 + \frac{\rho\,(\cos(\theta_0)\cos(\alpha)-\sin(\theta_0)\sin(\alpha))}{\rho_0(\rho_0-\rho)}, \quad (4)$$

$$u_1 = \frac{\rho_0 \sin(\theta_0)\cos(\alpha)+\rho\cos(\theta_0)\sin(\alpha)}{\rho_0 d_1(\rho_0-\rho)+\rho(\cos(\theta_0)\cos(\alpha)-\sin(\theta_0)\sin(\alpha))}. \quad (5)$$

Analyzing the linearity index proposed in [7]:

$$L_\rho = \left| \frac{\frac{\partial^2 u}{\partial \rho^2}\Big|_{\rho=0} 2\sigma_\rho}{\frac{\partial u}{\partial \rho}\Big|_{\rho=0}} \right|, \quad (6)$$

we obtain:

$$L_\rho = \frac{2\sigma_\rho}{\rho_0} 2 \left| 1 - \frac{d_0}{d_1}(cos(\theta_0)cos(\alpha) - sin(\theta_0)sin(\alpha)) \right|. \quad (7)$$

In order to have an acceptable linearization, it should be $L_\rho \approx 0$. The result obtained shows that, in this analysis, we can not ignore the $\theta_0$ if $\rho_0 \approx 4\sigma_\rho$, and, when we have a low parallax angle, $\theta_0$ becomes important since the term $(1-\frac{d_0}{d_1}(\cos(\theta_0)\cos(\alpha)-\sin(\theta_0)\sin(\alpha))) \rightarrow (1-\frac{d_0}{d_1}\cos(\theta_0))$. In Figure 2 it is possible to see the value of this term as the parallax angle increases. Another important linearity consideration concerns the initialization procedure: every time the camera perceives a new feature, we have to estimate the value of its $\theta$ angle (and, in 3D, also the value of $\phi$). This

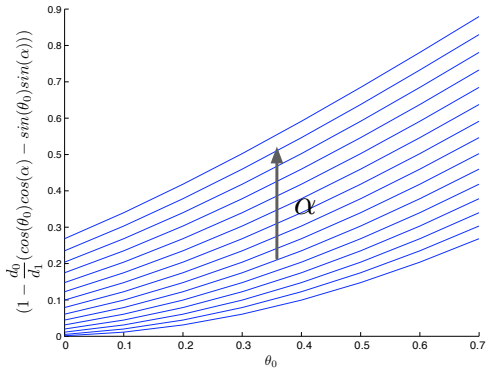



Fig. 2. UID parametrization linearity analysis: value of term $(1-\frac{d_0}{d_1}(\cos(\theta_0)\cos(\alpha)-\sin(\theta_0)\sin(\alpha)))$ when $\alpha$ goes from 0.05 to $\pi/4$, $\theta_0$ from 0 to $\pi/4$ and $d_0 \approx d_1$.

introduces another non-linearity factor in the measurement equation, since this has to be taken into account in the Jacobian calculation. These observations motivate our work. We propose to change the parametrization, in order to avoid this coordinate transformation and to reduce further the non-linearity of the measurement equation. Our idea, introduced in its preliminary form in [8], is based on the observation that it is possible to represent a 3D point in the scene by appropriately scaling the triangle formed by the image point, the image center, and projection center, which turns into using homogenous coordinates:

$$\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = \frac{1}{\omega_i} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix}. \quad (8)$$

Taking into account the inverse scaling $1/\omega_i$, we can model the uncertainty skewness as with the UID approach and, at the same time, simplify the measurement equation. Following this intuition we can rewrite the formulae for the error propagation analysis (see again Figure 1 for a reference):

$$D = \frac{1}{\omega_0 - \omega}, \quad d_0 = \frac{1}{\omega_0}, \quad (9)$$

$$d = D - d_0 = \frac{\omega}{\omega_0(\omega_0-\omega)}. \quad (10)$$

The image point in the second camera will be:

$$x_1 = \frac{u\cos(\alpha)\omega_0 + \omega\sin(\alpha))}{\omega_0(\omega_0-\omega))}, \quad (11)$$

$$y_1 = d_1 + \frac{\omega\cos(\alpha) - u_0\sin(\alpha)}{\omega_0(\omega_0-\omega)}, \quad (12)$$

$$u_1 = \frac{u_0\cos(\alpha)\omega_0 + \omega\sin(\alpha)}{d_1\omega_0(\omega_0-\omega)+\omega(\cos(\alpha)-u_0\sin(\alpha))}, \quad (13)$$

and the linearity index $L_\omega$:

$$L_\omega = \left| \frac{\frac{\partial^2 u}{\partial \omega^2}\Big|_{\omega=0} 2\sigma_\omega}{\frac{\partial u}{\partial \omega}\Big|_{\omega=0}} \right| = \quad (14)$$

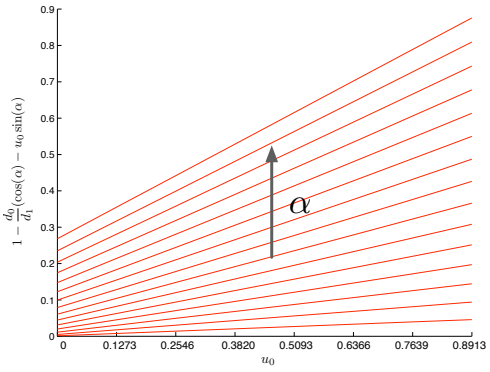



Fig. 3. Inverse scaling linearity analysis: value of term $(1-\frac{d_0}{d_1}(\cos(\alpha)-u_0\sin(\alpha))$ when $\alpha$ goes from 0.05 to $\pi/4$, $u_0$ from 0 to 1 (this range corresponds, for a normalized camera, to the range of $\theta_0$ in Figure 2: we calculate the value of $u_0$ for each $\theta_0$ thus it is possible to have a direct correspondence between the two graphs) and $d_0 \approx d_1$.

$$= \frac{2\sigma_\omega}{\omega_0} 2 \left| 1 - \frac{d_0}{d_1}(\cos(\alpha) - u_0 \sin(\alpha)) \right|. \quad (15)$$

In this case, when we have a low parallax ($\alpha \to 0$, $\cos(\alpha) \to 1$ and $\sin(\alpha) \to 0$), the displacement $u_0$ will be balanced by the value of $\sin(\alpha)$, so improving the equation linearity as clearly stated from Figure 3. These results are confirmed by experiments with simulated data, presented in Section IV.

## III. MONOSLAM USING INVERSE SCALING

The parametrization proposed in the previous section has been validated as part of a complete SLAM system that uses an Extended Kalman Filter to jointly represent the map of the world and the robot pose. In this paper, we consider the camera pose represented by six degrees of freedom, and a sensor providing 2D data.

State representation in a EKF-based SLAM system is:

$$\mathbf{x} = \left[ \begin{array}{cccccc} \mathbf{x}_R^W & \mathbf{v}^R & \mathbf{x}_{F_1}^W & \dots & \mathbf{x}_{F_m}^W & \dots & \mathbf{x}_{F_M}^W \end{array} \right]^T \quad (16)$$

being $\mathbf{x}_R^W = [\phi, \gamma, \theta, x, y, z]^T$ the six degrees of freedom representation of the camera pose w.r.t. the world reference frame $W$, $\mathbf{v}^R = [v_\phi, v_\gamma, v_\theta, v_x, v_y, v_z]^T$ is the camera velocity w.r.t the camera pose, and $\mathbf{x}_{F_m}^W = [x, y, z, \omega]^T$ is the Inverse Scaling parametrization of the feature w.r.t. the world reference frame $W$.

A constant linear and angular velocity is assumed and this produces, at each step, a roto-translation $\mathbf{x}_{R_k}^{R_{k-1}}$ between the previous camera reference system ($R_{k-1}$) and the actual pose ($R_k$). Moreover, at each step we assume an additive white and zero mean Gaussian error due to an unknown acceleration factor $\mathbf{a}$ with covariance $\mathbf{Q}$.

$$\mathbf{v}^{R_{k-1}} = \hat{\mathbf{v}}^{R_{k-1}} + \mathbf{a} \cdot \Delta t, \quad (17)$$

$$\mathbf{x}_{R_k}^{R_{k-1}} = \hat{\mathbf{v}}^{R_{k-1}} \cdot \Delta t. \quad (18)$$

The state is updated in two steps: *prediction* and *update*. The state, after the prediction step will be:

$$\mathbf{x}_{k|k-1} = \left[ \begin{array}{c} \mathbf{x}_{R_{k-1}}^W \oplus \mathbf{x}_{R_k}^{R_{k-1}} \\ \mathbf{v}^{R_k} \\ \mathbf{x}_{F_1}^W \\ \vdots \\ \mathbf{x}_{F_m}^W \end{array} \right], \quad (19)$$

where: $\mathbf{v}^{R_{k-1}} = \mathbf{v}^{R_{k-1}} + \mathbf{a}_k \Delta t, \mathbf{x}_{R_k}^{R_{k-1}} = \mathbf{v}^{R_{k-1}} \Delta t, \mathbf{v}^{R_k} = \mathbf{v}^{R_{k-1}}$; $\oplus$ is the transformation composition operator. The corresponding covariance is:

$$\mathbf{P}_{k|k-1} = \mathbf{J}_1 \mathbf{P}_{k-1|k-1} \mathbf{J}_1^T + \mathbf{J}_2 \mathbf{Q} \mathbf{J}_2^T \quad (20)$$

being

$$\mathbf{J}_1 = \left[ \begin{array}{cccc} \mathbf{J}_x & \mathbf{J}_v & \dots & \mathbf{J}_{F_m} \end{array} \right], \quad \mathbf{J}_2 = \left[ \begin{array}{c} \mathbf{J}_{a_k} \end{array} \right] \quad (21)$$

with

$$\mathbf{J}_x = \frac{\partial \mathbf{x}_{k|k-1}}{\partial \mathbf{x}_{R_{k-1}}^W}, \mathbf{J}_v = \frac{\partial \mathbf{x}_{k|k-1}}{\partial \mathbf{v}_{R_k}^{R_{k-1}}}, \quad (22)$$

$$\mathbf{J}_{F_m} = \frac{\partial \mathbf{x}_{k|k-1}}{\partial \mathbf{x}_{F_m}^W}, \mathbf{J}_{a_k} = \frac{\partial \mathbf{x}_{k|k-1}}{\partial \mathbf{a}_k}. \quad (23)$$

In order to define the measurement equation, one can immediately derive, from our parametrization, the following.

$$\mathbf{h}^{R_k} = \left[ \begin{array}{c} h_x^{R_k} \\ h_y^{R_k} \\ h_z^{R_k} \end{array} \right] = \mathbf{M} \cdot \mathbf{R}_W^{R_k} \left( \left[ \begin{array}{c} x_{F_i}^W \\ y_{F_i}^W \\ z_{F_i}^W \end{array} \right] - \omega_{F_i}^W \mathbf{r}_{R_k}^W \right) \quad (24)$$

where $\mathbf{h}^{R_k}$ is in homogeneous coordinates, $\mathbf{R}_W^{R_k}$ is the rotation matrix between robot pose at time $k$ and the world reference frame $W$, $\mathbf{r}_{R_k}^W$ is the translation vector between $W$ world reference frame and the robot pose at time $k$, $\mathbf{M}$ is the calibrated projection matrix of the camera, and $\mathbf{D}$ its covariance:

$$\mathbf{M} = \left[ \begin{array}{ccc} fc_x & 0 & cc_x \\ 0 & fc_y & cc_y \\ 0 & 0 & 1 \end{array} \right], \quad (25)$$

$$\mathbf{D} = \left[ \begin{array}{cccc} \sigma_{fc_x}^2 & 0 & 0 & 0 \\ 0 & \sigma_{fc_y}^2 & 0 & 0 \\ 0 & 0 & \sigma_{cc_x}^2 & 0 \\ 0 & 0 & 0 & \sigma_{cc_y}^2 \end{array} \right]; \quad (26)$$

$\mathbf{x}_W^{R_k}$ is the roto-translation matrix between pose $k$ and the world reference frame $W$; $\mathbf{h}^{R_k}$ above is the projection of the 3D point in the camera frame, in homogeneous coordinates. The measurements, i.e., the pixel coordinates, are:

$$\mathbf{h}_k = \left[ \begin{array}{c} h_{k_u} \\ h_{k_v} \end{array} \right] = \left[ \begin{array}{cc} \frac{h_x^{R_k}}{h_z^{R_k}}, \frac{h_y^{R_k}}{h_z^{R_k}} \end{array} \right]^T. \quad (27)$$

Moreover, we add $\mathbf{D}$ to the state covariance matrix $\mathbf{P}_{k|k-1}$:

$$\mathbf{P}_{k|k-1} = \left[ \begin{array}{cc} \mathbf{P}_{k|k-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{array} \right]. \quad (28)$$

In this way we take the uncertainty of the projection matrix into consideration.

The classical EKF update equations give the new estimate of both the state vector $\mathbf{x}_{k|k}$ and the camera motion from the world reference frame to the camera pose $k$.

$$\begin{aligned} \mathbf{S} &= \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k \\ \mathbf{K} &= \mathbf{P}_{k|k-1} \mathbf{H}_k^T \mathbf{S}^{-1} \\ \mathbf{P}_{k|k} &= \mathbf{P}_{k|k-1} - \mathbf{K}\mathbf{S}\mathbf{K}^T \\ \mathbf{x}_{k|k} &= \mathbf{x}_{k|k-1} + \mathbf{K}\left(\mathbf{z} - \mathbf{h}_k\right) \end{aligned} \tag{29}$$

where $\mathbf{R}_k$ is the measurement error covariance, $\mathbf{z}$ the observations and $\mathbf{H}_k$:

$$\mathbf{H}_k = \frac{\partial \mathbf{h}_k}{\partial \mathbf{x}_{k-1}} = \left[ \mathbf{H}_{x_{R_k}^W} \; \mathbf{H}_{v_k} \; \mathbf{H}_{F_1} \; \ldots \; \mathbf{H}_{F_i} \; \ldots \; \mathbf{H}_{F_m} \; \mathbf{H}_M \right], \tag{30}$$

where

$\mathbf{H}_{x_{R_k}^W} = \frac{\partial \mathbf{h}_k}{\partial \mathbf{x}_{R_k}^W}$, $\mathbf{H}_{v_k} = \frac{\partial \mathbf{h}_k}{\partial \mathbf{v}^{R_k}}$, $\mathbf{H}_{F_i} = \frac{\partial \mathbf{h}_k}{\partial \mathbf{x}_{F_i}^W}$, $\mathbf{H}_M = \frac{\partial \mathbf{h}_k}{\partial \mathbf{M}}$.

### A. Initialization of a new feature

The initialization of a new feature, though perceived from the camera, has to be moved in the map reference frame, unless to follow a robo-centric approach, like it was done in [8]. With Inverse Scaling, we can initialize the features with a huge uncertainty in the depth, as with UID, in order to represent that most of the uncertainty lays in the direction of the interpretation ray. Moreover, the uncertainty of the map is described by Gaussian distributions over the parameters, in Inverse Scaling, as it was with UID.

Each feature in the camera reference frame is defined as:

$$\mathbf{x}_{F_{new}}^{R_k} = \left( \; x, y, z, \omega \; \right)^T; \tag{31}$$

when we obtain an observation $h = (u, v)^T$ of a new feature from the camera, we initialize its parameters as:

$$\begin{pmatrix} x \\ y \\ z \\ \omega \end{pmatrix} = \begin{pmatrix} u - cc_x \\ v - cc_y \\ fc \\ \hat{\omega} \end{pmatrix}; \tag{32}$$

being $fc$ the focal length of the camera (we suppose unit aspect ratio), $[u, v]$ the 2D image point and $[cc_x, cc_y]$ the projection center of the camera. The initial value of $\hat{\omega}$ can be defined so to cover the entire working range; for $1/\omega$ uncertainty to cover (with 96% probability) the range between some minimum distance $min_d$ to infinite, $\omega$ needs to be in the 4% confidence interval $[0, 1/min_d]$. In our experiments, we used initial $\hat{\omega} = fc/(2 * min_d)$ and $\sigma_\omega = fc/(4 * min_d)$; this allows to represent the non-normal uncertainty extending from $min_d$ to infinite.

Feature initialization takes into consideration all information available without any linearization process. The projection matrix values and the observations are taken as initial values for the inverse scaling variables. The uncertainty on projection matrix and on observations are used to initialize the uncertainty on these variables in linear way. The result is that we have inverse scaling parameters represented correctly by Gaussian stochastic variables and we can represent a point to infinity without any special trick.

Subsequently, we have to roto-translate this new feature in the world coordinate frame, but still in homogeneous form.

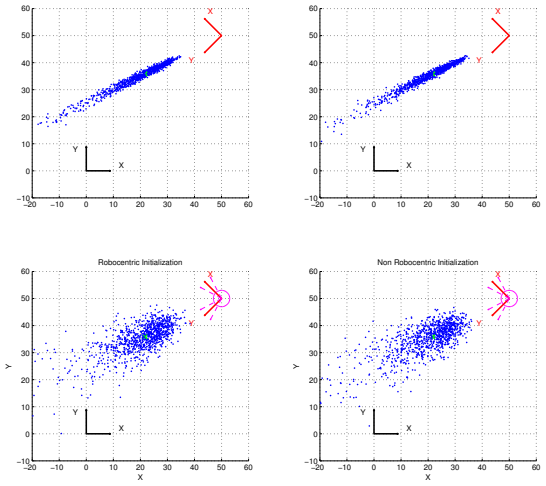



Fig. 4. In the first row we report the samples from the uncertainty distribution of a feature in the camera frame of reference (on the left) and in the world reference frame after roto-translation. In the second row we introduced roto-translation uncertainty (represented in the plot by the $3\sigma$ angles and uncertainty ellipse) in the camera reference frame initialization (left) and after Jacobian uncertainty propagation (right).

The new state covariance, after the initialization, is obtained using the image measurement error covariance $\mathbf{R}_k$, the state vector covariance $\mathbf{P}_{k|k}$, and the projection matrix covariance $\mathbf{D}$ (to keep in consideration the uncertainty on the camera parameters). It becomes:

$$\mathbf{x}_{k|k} = \begin{bmatrix} \mathbf{x}_{k|k} \\ \mathbf{x}_{R_k}^W \oplus \mathbf{x}_{F_{new}}^{R_k} \end{bmatrix} \tag{33}$$

$$\mathbf{P}_{k|k} = \mathbf{J} \begin{bmatrix} \mathbf{P}_{k|k} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_k & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & {\sigma_\omega}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D} \end{bmatrix} \mathbf{J}^T \tag{34}$$

with:

$$\mathbf{J} = \left[ \begin{array}{c|cccc} \mathbf{I} & \mathbf{0} & & & \\ \hline \frac{\partial \mathbf{x}_{F_{new}}^W}{\partial \mathbf{x}_{R_k}^W}, \mathbf{0} & \frac{\partial \mathbf{x}_{F_{new}}^W}{\partial h} & \frac{\partial \mathbf{x}_{F_{new}}^W}{\partial \omega} & \frac{\partial \mathbf{x}_{F_{new}}^W}{\partial M} \end{array} \right]. \tag{35}$$

In a previous work [8] we applied a robo-centric approach so this rotation was not needed in the initialization phase; it might be objected that by the application of this rotation we loose the skewness of the distribution and, with it, the effectiveness of our novel parametrization. It turns out that this roto-translation is applied to the inverse scaling parametrization to obtain a new inverse scaling parametrization in the world frame of reference (i.e., roto-translation happens in the homogeneous coordinates space), and thus the proper modeling of uncertainty is preserved. To show this we simulated this operation and the result is reported in Figure 4. The two plots in the upper part of the figure represent the feature initialized in the camera frame (left) and in the world reference frame (right). Samples in the camera frame (top left in Figure 4) are generated by sampling the feature

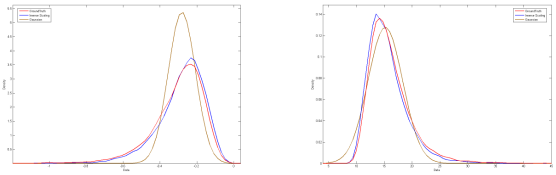


Fig. 5. Belief distribution for a 2D scene point 15m away from the observer: (red) true distribution (computed with particle transformation), (blue) Inverse Scaling parametrization, (brown) classical parametrization (Gaussian approximation via Jacobians). $x$ is on the left, $y$ on the right.

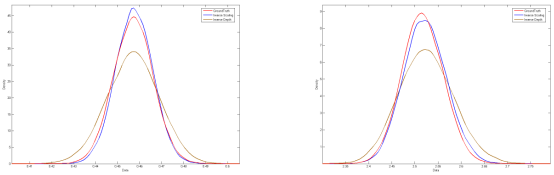


Fig. 6. Belief distribution for a 2D scene point 2.5m away from the observer: (red) true distribution (computed with particle transformation), (blue) Inverse Scaling parametrization, (brown) UID parametrization. $x$ is on the left, $y$ on the right.

initialized in the camera frame; each sample is then roto-translated in the world reference frame by applying the exact roto-translation between the two frames. The samples in the world reference frame (top right in Figure 4) are directly sampled from the inverse scaling representation after the roto-translation applied to it as from Equations 33, and 34. When adding uncertainty to the roto-translation we obtain the plots in the second row of Figure 4; the samples on the left are from the feature initialization, and are then combined with a sample from the roto-translation distribution; the samples on the right are sampled directly from the feature distribution after the uncertainty propagation. It is clear from the simulations that the skewed uncertainty is preserved and the deformation introduced by the linearization process is negligible.

## IV. EXPERIMENTAL RESULTS

In this section we present the capabilities of our representation using a simulator for a monocular vision system. Given a point in the map, and the position of the camera w.r.t. the map, we simulate the image formation on the device, as well as the uncertainty of the measurements. The motivation for using a simulated environment to test the proposed model is to have access to the ground truth and therefore to compare different methods using the same data. On the other hand, in simulation we can easily use a sample based approach to produce a proper representation of the true uncertainty through exact particle triangulation as we did in the previous section.

The simulated world is planar with 1D cameras, this totally suffice to prove the paper claims. The parameters used for the simulated monocular system are: image resolution of 640 pixels at 30Hz and an uncertainty associated to the image measurements set to $\sigma$=0.5 pixels. We consider the projection matrices known altogether with their uncertainty, assumed normal; focal length of 650 pixels with an uncertainty of $\sigma$=3 pixels and projection center of 320 pixel with $\sigma$=2 pixels. The purpose of the experiments is to analyze the performance of the inverse scaling parametrization with features at different locations and depths.

The graphs in Figure 5, 6, and 7 represent the Probability Distribution Function along the axis. In Figure 5 it is possible to compare the triangulation result using our model with the classical approach, i.e., Jacobian uncertainty propagation and $[x, y, z]^T$ point representation. The graph shows the

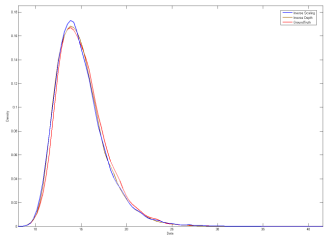


Fig. 7. Belief distribution for a 2D scene point 15m away from the observer: (red) true distribution (computed with particle transformation), (blue) Inverse Scaling parametrization, (brown) UID parametrization. Only the $y$ coordinate is depicted.

reconstruction of a scene point at 15m from the observer using stereo cameras with baseline of 0.6m. We can see the non-gaussian-ness of the real distribution in comparison with the classical Gaussian representation. Moreover, notice the better distribution approximation of inverse scaling.

In Figure 6 we compare the uncertainty distribution generated using Inverse Scaling versus the UID approach when we try to estimate the 2D scene point at 2.5m (i.e., with a large parallax angle). The plot shows that the distribution estimated by our model is realistic in different experimental condition, i.e., with both the large and the small parallax conditions. This property allows to use our representation in different real conditions. Figure 7 shows that the two parameterizations are more or less equivalent for long range data.

Finally, in figure 8 we shown the distribution, coded with the classical cartesian point representation $(X, Y)$, of the same point. Notice the non-Gaussian distribution of these variables, after triangulation. Figure 9 shows the uncertainty on a 2D point coded with the inverse scaling parametrization. It remains Gaussian after triangulation.

### A. Comparison within SLAM

To verify if a better uncertainty modeling lead to better SLAM results (somehow confirming the results in [9]), we tested two SLAM systems in a simulated rectangular environment (point features are equally distributed along the environment borders); the former implements what is proposed in Section III, the latter uses the UID parametrization. Our UID implementation is an adapted version of [3] to the particular case of 3DoF simulation (matrixes instead of

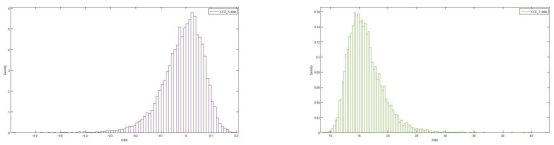


Fig. 8. Estimated distribution of a 2D scene point 15m away from the observer coded with cartesian $(X, Y)$ representation. It is computed with a particle transformation through the triangulation process. The $x$ coordinate is depicted above, the $y$ below.

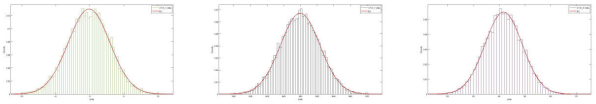


Fig. 9. Estimated distribution of the same 2D scene point as above when coded with inverse scaling $(X, Y, W)$ representation. The $x$ coordinate is depicted above, the $y$ and $w$ below, respectively.

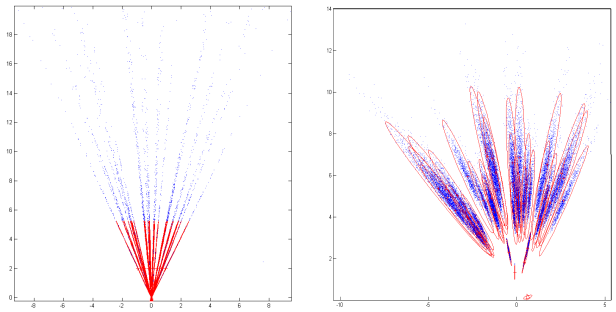
quaternions for rotations). In simulation the data association has been performed manually, so that estimates are comparable and the main aspect considered is uncertainty modeling. We assume obvious to the reader, for the evaluation of the simulation results, that the differences between the two approaches in a real setting, i.e., without relying on a correct data association, can only be larger.

Figure 10(a) represents the initialization process. This figure shows the comparison between a cartesian Gaussian parametrization of the features (in red) and our inverse scaling Gaussian parametrization (represented by the cloud of points in blue). The standard Gaussian uncertainty ellipsoids are obtained by a Jacobian uncertainty propagation (linearization) to transform the 3 dimensional ($x$,$y$,$\omega$) inverse scaling coordinates in 2 dimensional cartesian ($x$,$y$) coordinates. The inverse scaling representation is obtained by sampling directly from the 3 dimensional ellipsoid ($x$,$y$,$\omega$).

We notice that at the beginning we have a huge depth uncertainty (from zero to infinity) as we do not have any information about the depth of the features. This situation is easily coded in the inverse scaling parametrization through the $\omega$ factor. This allows to represent features seen for the first time (where we know the direction of the pixel interpretation ray, but we do not know exactly where the features is along this ray). After the initialization step, the camera moves. This movement produces parallax and thereby the features depth estimate is improved, reducing the uncertainty on the features (see Figure 10(b)).

In Figure 11, we have the plots of the error in pose estimation, respectively for $x$, $y$ and $\theta$, during the robot path. The path is a simple circle with the camera always looking outside, i.e., toward the borders of the environment. As it can be easily noticed the variance of the robot pose estimate (the blue lines are placed at $\pm 3\sigma$) is underestimated for the UID parametrization; this is not the case for the Inverse Scaling parametrization. The underestimation leads to filter inconsistency.

Finally, we present a real application of our system in



(a) (b)

Fig. 10. (a) Initialization: in blue the particles representing the uncertainty coded with Inverse Scaling, in red a cartesian ($x$,$y$) Gaussian representation of the uncertainty coded by using the Jacobian uncertainty propagation (notice that there is no red in y beyond 5m, for the sake of those looking at a grey-level version of the picture). (b): after 500 steps. Notice that the distributions reached equivalence to normality.

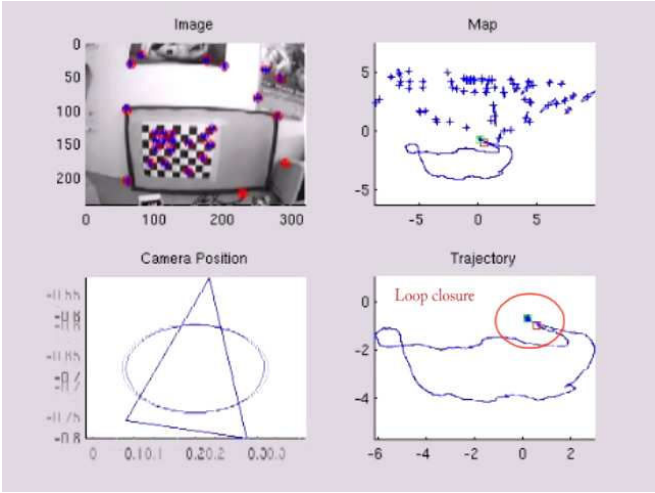



Fig. 12. Map reconstruction using Inverse Scaling parametrization in a real indoor environment. Images are from the video attached to the paper.

a real indoor context. Data association is here based on patch correlation. Also, not all the detectable features are used, a sub-sampling scheme is adopted to reduce their number. In Figure 12 there are some frames taken using a 320x240 B/W camera at 30Hz, from the video attached to this paper. The camera was hand held and moved. The figure shows the results of the estimation process using the proposed monocular approach: the top-left image shows the camera image (here we have in red the prediction for the features, and in blue the matched ones); the top-right image presents the estimated map, from the top, with the uncertainty ellipsoids, the bottom-left shows the camera position with its uncertainty ellipsoid, and the bottom-right features the camera trajectory, from the top.

## V. CONCLUSIONS AND DISCUSSION

In this paper we introduced a new parametrization, called *inverse scaling*, for monocular SLAM based on EKF filter. Compared with the UID solution [3], our approach allows to improve the accuracy of the uncertainty modeling, simplifies

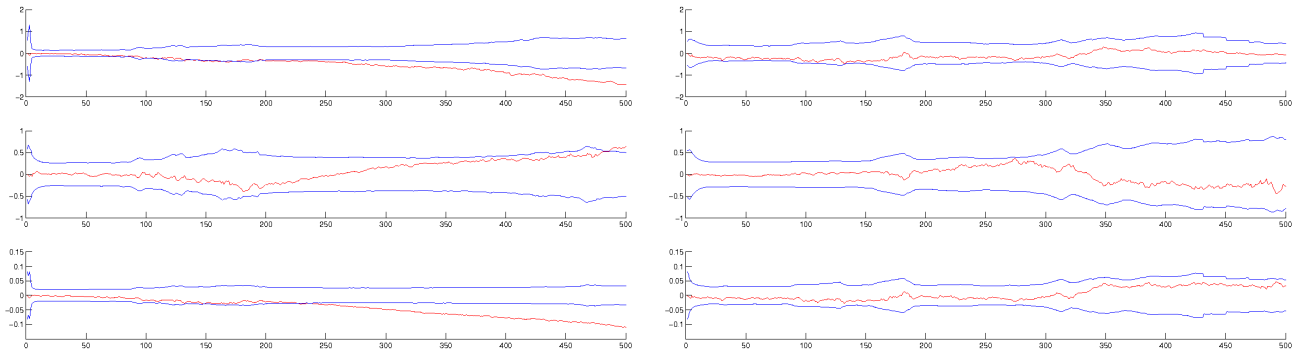


Fig. 11. Error in robot localization ($x$,$y$,$\theta$): (left) using UID parametrization, (right) using Inverse Scaling parametrization. In red the error w.r.t. the ground truth, in blue $\pm 3\sigma$

the measurement equation, and reduces its non-linearity.

It might be argued that inverse scaling parametrization is somehow over-parameterized (and this apply even more for the UID parametrization) since we use 4 parameters to represent a 3D point that could be represented with the classical 3 Euclidean coordinates. Our claim is that, by using the extra parameter and representing the point in the space of homogeneous coordinates, we achieve the following goals:

- make the measurement equation linear (or close to linear) as previously shown (see Section II);
- properly initialize a new feature with only a single view of it (see Section III-A);
- compute the uncertainty on this feature tacking into account both the image and the projection uncertainty.
- represent the uncertainty of this features, skewed in the Euclidean space, as Gaussian in the space of homogeneous coordinates (see Section IV);
- achieve all these claims using only 4 parameters (instead of the 6 UID parameters).

Lastly, our parametrization makes feature initialization and measurement linear or close to linear also w.r.t. the projection parameters; therefore it is possible to consider the projection uncertainty in both the measurement and the initialization (see Section III-A).

We validated our claims both mathematically, extending the study done by Civera et al. [7], and experimentally, using both a simulated framework, to allow comparison with ground truth and a real setup.

## VI. ACKNOWLEDGMENTS

This work has been partially supported by the European Commission, Sixth Framework Programme, Information Society Technologies: Contract Number FP6-045144 (RAWSEEDS), and by an Italian Institute of Technology (IIT) grant. We thank Prof. J.D. Tardos for his comments on a draft of the paper, which helped us improving it.

## REFERENCES

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

[2] J. Solà, A. Monin, M. Devy, and T. Lemaire, "Undelayed initialization in bearing only slam," in *IEEE International Conference on Intelligent Robots and Systems, IROS 2005.*, 2005, pp. 2499–2504.

[3] J. Montiel, J. Civera, and A. Davison, "Unified inverse depth parametrization for monocular slam," in *Proceedings of Robotics: Science and Systems*, 2006 August.

[4] E. Eade and T. Drummond, "Scalable monocular slam," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 469–476, 2006.

[5] ——, "Monocular slam as a graph of coalesced observations," in *Proceedings of IEEE International Conference on Computer Vision 2007*, 2007 October.

[6] L. A. Clemente, A. J. Davison, I. D. Reid, J. Neira, and J. D. Tardos., "Mapping large loops with a single hand-held camera." in *In Proceedings of Robotics: Science and Systems*, 2007.

[7] J. Civera, A. J. Davison, and J. M. M. Montiel, "Inverse depth to depth conversion for monocular slam," in *ICRA*, 2007, pp. 2778–2783.

[8] D. Marzorati, M. Matteucci, D. Migliore, and D. G. Sorrenti, "Monocular slam with inverse scaling parametrization," in *Proceedings of 2008 British Machine Vision Conference (BMVC 2008)*, R. F. M. Everingham, C.J. Needham, Ed., 2008, pp. 945–954.

[9] D. Marzorati, M. Matteucci, and D. G. Sorrenti, "Particle-based sensor modeling for 3d-vision slam," in *Proceedings of IEEE International Conference on Robotics and Automation 2007*, April 2007.