Time Shifted IMU Preintegration for Temporal Calibration in Incremental Visual-Inertial Initialization

Bruno Petit, Richard Guillemard and Vincent Gay-Bellile
Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France
bruno.petit@ensta.org, {richard.guillemard,vincent.gay-bellile}@cea.fr

Abstract

Tightly coupled Visual-Inertial SLAM (VISLAM) algorithms are now state of the art approaches for indoor localization. There are many implementations of VISLAM, like filter-based and non-linear optimization based algorithms. They all require an accurate temporal alignment between sensors clock and an initial IMU state (gyroscope and accelerometer biases value, gravity direction and initial velocity) for precise localization. In this paper we propose an initialization procedure of VISLAM that estimates simultaneously IMU-camera temporal calibration and the initial IMU state. To this end, the concept of Time Shifted IMU Preintegration (TSIP) measurements is introduced: an interpolation of IMU preintegration that takes into account the effect of sensors clock misalignment. These TSIP measurements are included along with visual odometry measurements in a graph that is incrementally optimized. It results in a real time, accurate and robust initialization for VISLAM as demonstrated in the experiments on real data.

1. Introduction

Visual-inertial odometry is an active research topic in robotics and computer vision communities. This is the most widespread and efficient solution to solve indoor localization problem. Cameras and Inertial Measurement Units (IMU) are small, cheap and low power consumption sensors. IMU brings robustness by giving a motion prediction, and visual information ensures precision of the localization by reducing the divergence of cheap IMU-only localization and allowing to observe IMU biases. VISLAM algorithms are used as dead reckoning tools and/or to build metric scale dense maps. Given an initial state, position and orientation are estimated in real time using tight sensor fusion. Implementations of VISLAM can be split in two main families. Filtered-based [10, 16] algorithms use an extended Kalman filter on a multi state estimation (MSCKF). In contrast, non-linear optimization methods [12, 6] fuse visual and inertial data in a bundle adjustment where inertial constraints are added through the concept of IMU preintegration [6].

Regardless of the visual-inertial fusion algorithm used, the localization accuracy they reach greatly relies on temporal alignment of sensor streams [5, 13]. Since high precision sensor timestamping is expensive and requires a high hardware expertise, streams are often software timestamped. It leads to a time shift between each sensor data sampling instant and its attached timestamp. Hardware sensors triggering allows to prevent stream temporal drift but a constant time delay between sensors data timestamps remains. The value of this delay sometimes differs too significantly from one acquisition to another and therefore often needs to be estimated at every use of the system.

State of the art IMU-camera time shift estimations are based on the fact that these sensors are able to measure redundant information. For example the common tool Kalibr [14] estimate temporal calibration and IMU biases by interpolation and correlation of visual odometry trajectory with IMU motion prediction. Nevertheless, althought offering great precision and robustness, this method suffers from being too computation costly to be executed in real time. Recent works have been focusing on online time shift estimation for visual-inertial systems. Several methods propose a temporal calibration incorporated in a VISLAM, supposedly well initialized beforehand. In [13] a time shift estimation in a bundle-based VISLAM is proposed. It is build on visual feature interpolation by introducing the concept of feature velocity that enables to predict the time shift impact on feature 2D position. In [18] a filtered-based VISLAM is used to estimate time shift and its observability is studied. Online methods work sequentially, starting by estimating IMU initial state in order to run a VISLAM with time shift estimation in a second phase. However, without temporal alignment, the initialization of the IMU state may be inaccurate leading to VISLAM failure as shown in tests carried out in [5].

Both time shift and initial IMU state have to be simultaneously initialized for accurate VISLAM localization. To this end we propose in this paper two contributions.

The first contribution of this work is the concept of Time Shifted IMU Preintegration (TSIP) measurements: a temporal interpolation of IMU preintegration measurements [6]. To the best of our knowledge, no method that takes into account time shift in preintegration measurements of raw IMU data has been proposed before. TSIP are designed to be efficiently estimated in a non-linear optimization algorithm and are therefore well suited for real time use.

The second contribution is an incremental probabilistic estimation of both time shift and initial visual-inertial state in a single non-linear optimization. The proposed TSIP measurements and poses from visual odometry are taken as input. The robustness of the proposed initialization is increased by taking into account the uncertainty of the visual odometry and by handling odometry failures.

After introducing some preliminaries in section 2, the concept of time shifted preintegration measurement is presented in section 3 and the non-linear optimization for IMU initialization and time shift estimation is detailed in section 4. Experimental evaluations are presented in section 5.

2. Preliminaries

From camera and IMU sensors, visual-inertial systems provide two types of data streams:

- A visual stream, bringing 2D features that are detected then matched in images of calibrated camera(s),
- An inertial stream, that includes synchonized gyroscope and accelerometer data.

These streams of data are preprocessed before being used in the proposed initialization algorithm: the matched features are used to estimate an odometry and its uncertainty, and IMU raw data are preintegrated.

2.1. Visual odometry

A Visual SLAM [11] is used to create keyframe poses from the video stream in real time. They are rigid transformations from a body frame B to a world frame W at instants t_i:

$$P_{\mathrm{WB}}(t_i) = \begin{bmatrix} R_{\mathrm{WB}}(t_i) & {}_{\mathrm{W}}\mathbf{p}_{\mathrm{B}}(t_i) \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \in SE(3). \tag{1}$$

For clarity of notation, frames indication "WB" for odometry poses is dropped in future expressions, *i.e.* $P_{\mathrm{WB}}(t_i) = P_i$, $R_{\mathrm{WB}}(t_i) = R_i$ and ${}_{\mathrm{W}}\mathbf{p}_{\mathrm{B}}(t_i) = \mathbf{p}_i$. These poses are optimized in a bundle adjustment and gathered in the set $\mathcal{P} \doteq \{P_i\}_{i=1\dots N}$.

To evaluate the quality of the localization, odometry uncertainty $\Sigma^{\mathcal{P}}$ is recovered from the information matrix of the bundle adjustment. In practice, the costly computation of the entire covariance matrix $\Sigma^{\mathcal{P}}$ has to be avoided. Therefore, only parts of this matrix are computed by marginalization [9] to summerize the uncertainty of every delta pose $\Delta P_{i,i+1} = P_i^{-1} P_{i+1}$ in *inter pose covariances* $\Sigma^{\Delta P_{i,i+1}}$ defined as:

$$\Sigma^{\Delta P_{i,i+1}} = J_{i,i+1}\, \mathrm{Cov}(P_i, P_{i+1})\, J_{i,i+1}^{\top}, \tag{2}$$

$$J_{i,i+1} = \begin{bmatrix} R_{i+1}^{\top} R_i & \mathbf{0}_{3\times 3} & -\mathbf{I}_{3\times 3} & \mathbf{0}_{3\times 3} \\ \mathbf{0}_{3\times 3} & R_{i+1}^{\top} R_i & \mathbf{0}_{3\times 3} & -\mathbf{I}_{3\times 3} \end{bmatrix}. \tag{3}$$

with $\mathrm{Cov}(P_i, P_{i+1})$ being the covariance matrix of the Lie representation of P_i and P_{i+1}, obtained by marginalization.

2.2. IMU Preintegration

The IMU stream contains triaxial gyroscope and accelerometer data measuring the angular velocity ${}_{\mathrm{I}}\omega \in \mathbb{R}^3$ and the acceleration ${}_{\mathrm{I}}\mathbf{a} \in \mathbb{R}^3$ of the IMU frame I at regular intervals. Both measurements are affected by sensor noise η_g and η_a, and slow varying biases of the gyroscope and the accelerometer : $\mathbf{b}_{\mathrm{g}}$ and $\mathbf{b}_{\mathrm{a}}$. Moreover the accelerometer is subject to gravity ${}_{\mathrm{I}}\mathbf{g}$, leading to following measurements:

$$\tilde{\omega}(t) = {}_{\mathrm{I}}\omega(t) + \mathbf{b}_{\mathrm{g}}(t) + \eta_g(t) \tag{4a}$$

$$\tilde{\mathbf{a}}(t) = {}_{\mathrm{I}}\mathbf{a}(t) - {}_{\mathrm{I}}\mathbf{g}(t) + \mathbf{b}_{\mathrm{a}}(t) + \eta_a(t) \tag{4b}$$

In order to efficiently use IMU data in an optimization framework, raw gyroscope and accelerometer data are *preintegrated*. Motion between instants t_i and t_j is then defined in terms of this preintegration measurement to propagate the state of the IMU frame I:

$$R_{\mathrm{WI}}(t_j) = R_{\mathrm{WI}}(t_i)\, \Delta R_{i,j}, \tag{5a}$$

$${}_{\mathrm{W}}\mathbf{v}_{\mathrm{I}}(t_j) = {}_{\mathrm{W}}\mathbf{v}_{\mathrm{I}}(t_i) + {}_{\mathrm{W}}\mathbf{g}\, \Delta \mathrm{t}_{i,j} + R^i_{\mathrm{WI}} \Delta v_{i,j}, \tag{5b}$$

$$\begin{aligned} {}_{\mathrm{W}}\mathbf{p}_{\mathrm{I}}(t_j) = {}_{\mathrm{W}}\mathbf{p}_{\mathrm{I}}(t_i) + {}_{\mathrm{W}}\mathbf{v}_{\mathrm{I}}(t_i)\, \Delta \mathrm{t}_{i,j} \\ + \frac{1}{2}\, {}_{\mathrm{W}}\mathbf{g}\, \Delta \mathrm{t}_{i,j}^2 + R_{\mathrm{WI}}(t_i)\, \Delta p_{i,j}. \end{aligned} \tag{5c}$$

with preintegration measurement between instants t_i and t_j defined as:

$$\mathcal{I}_{i,j} \doteq \tag{6}$$

$$\begin{cases} \Delta R_{i,j} = \mathrm{Exp}\left(\int_{t_i}^{t_j} [\tilde{\omega}(t) - \mathbf{b}_{\mathrm{g}}(t)] dt \right), \\ \Delta v_{i,j} = \int_{t_i}^{t_j} \mathrm{Exp}\left(\int_{t_i}^{t} [\tilde{\omega}(s) - \mathbf{b}_{\mathrm{g}}(s)] ds \right) [\tilde{\mathbf{a}}(t) - \mathbf{b}_{\mathrm{a}}(t)] dt, \\ \Delta p_{i,j} = \iint_{t_i}^{t_j} \mathrm{Exp}\left(\int_{t_i}^{t} [\tilde{\omega}(s) - \mathbf{b}_{\mathrm{g}}(s)] ds \right) [\tilde{\mathbf{a}}(t) - \mathbf{b}_{\mathrm{a}}(t)] dt^2, \end{cases}$$

with $\mathrm{Exp}(\phi)$ standing for $\exp(\phi_\times)$ and $\phi_\times$ for the skew-symmetric matrix of ϕ. Discrete computation formulae of both $\mathcal{I}_{i,j}$ and its covariance $\Sigma^{\mathcal{I}_{i,j}}$ are explicited in [6].

Considering system states at two consecutive time instants i and j, preintegration measurement allow to define inertial residual errors on position, rotation and speed.

Those residual errors are given by [6]:

$$\mathbf{r}_{i,j}^{\Delta R} = \mathrm{Log}\left(\Delta R_{i,j}^{\top} R_{\mathrm{BI}}^{\top} R_i^{\top} R_j R_{\mathrm{BI}}\right), \tag{7a}$$

$$\mathbf{r}_{i,j}^{\Delta v} = {}_{\mathrm{W}}\mathbf{v}_{\mathrm{I}}(t_j) - {}_{\mathrm{W}}\mathbf{v}_{\mathrm{I}}(t_i) - {}_{\mathrm{W}}\mathbf{g}\,\Delta \mathrm{t}_{i,j} - R_i R_{\mathrm{BI}} \Delta v_{i,j}, \tag{7b}$$

$$\mathbf{r}_{i,j}^{\Delta p} = s(\mathbf{p}_j - \mathbf{p}_i) + (R_j - R_i)\,{}_{\mathrm{B}}\mathbf{p}_{\mathrm{I}} - {}_{\mathrm{W}}\mathbf{v}_{\mathrm{I}}(t_i)\,\Delta \mathrm{t}_{i,j} - \frac{1}{2}\,{}_{\mathrm{W}}\mathbf{g}\,\Delta \mathrm{t}_{i,j}^2 - R_i R_{\mathrm{BI}} \Delta p_{i,j}, \tag{7c}$$

where IMU extrinsic calibration is represented by a rigid transformation $T_{\mathrm{BI}} = [R_{\mathrm{BI}} \mid {}_{\mathrm{B}}\mathbf{p}_{\mathrm{I}}] \in SE(3)$, and $s \in \mathbb{R}$ is a scale factor. Log is the reciprocal bijection of Exp in Lie group of rotations $SO(3)$, expressions of these functions are explicited in [6].

The IMU preintegration of equation 6 directly depends on IMU biases, which are refined in the VISLAM optimization. To prevent its recomputation from raw IMU data at each iteration, Foster et al. proposed in [6] a linearization of the preintegration measurement taking into account biases update $\delta\mathbf{b}_{\mathrm{g}}$ and $\delta\mathbf{b}_{\mathrm{a}}$:

$$\hat{\mathcal{I}}_{i,j}(\delta\mathbf{b}_{\mathrm{g}}, \delta\mathbf{b}_{\mathrm{a}}) = \begin{cases} \Delta\hat{R}_{i,j} = \Delta R_{i,j}\,\mathrm{Exp}(J^{\mathrm{g}}_{\Delta R}\,\delta\mathbf{b}_{\mathrm{g}}) \\ \Delta\hat{v}_{i,j} = \Delta v_{i,j} + J^{\mathrm{g}}_{\Delta v}\,\delta\mathbf{b}_{\mathrm{g}} + J^{\mathrm{a}}_{\Delta v}\,\delta\mathbf{b}_{\mathrm{a}} \\ \Delta\hat{p}_{i,j} = \Delta p_{i,j} + J^{\mathrm{g}}_{\Delta p}\,\delta\mathbf{b}_{\mathrm{g}} + J^{\mathrm{a}}_{\Delta p}\,\delta\mathbf{b}_{\mathrm{a}} \end{cases} \tag{8}$$

with $J^{\mathrm{g}}_{(.)}$ and $J^{\mathrm{a}}_{(.)}$ standing for the jacobians of preintegration measurement with regard to IMU biases.

3. Time shifted IMU preintegration

IMU preintegration measurements presented in the previous section are taking into account IMU biases, but not the effect of a time offset. In this section a new IMU preintegration framework called Time Shifted IMU preintegration (TSIP) is proposed. It takes into account both biases and time shift influence and is designed for non-linear optimization. We first study the time shift influence on IMU preintegration measurements and propose a model for their temporal interpolation. Then the method to compute TSIP measurement is presented.

3.1. Temporal interpolation for IMU preintegration

Considering preintegration measurements $\mathcal{I}_{i,j}$ between two instants t_i and t_j, we seek a method to compute the value of $\bar{\mathcal{I}}_{a,b}^{i,j}$, an approximation of this preintegration measurement for any time instants t_a and t_b with $[t_a, t_b] \subset [t_i, t_j]$ without using raw IMU data.

To do so, gyroscope and accelerometer data as well as IMU biases are assumed to be constant during the preintegration time $\Delta \mathrm{t}_{i,j}$. Rewriting equation 6 with $\forall\, t \in [t_i, t_j]$, $\tilde{\omega}(t)\text{-}\,\mathbf{b}_{\mathrm{g}}(t) = \bar{\omega}_{i,j}$ and $\tilde{\mathbf{a}}(t)\text{-}\,\mathbf{b}_{\mathrm{a}}(t) = \bar{\mathbf{a}}_{i,j}$, gives:

$$\Delta R_{i,j} = \mathrm{Exp}(\bar{\omega}_{i,j}\Delta \mathrm{t}_{i,j}) \tag{9a}$$

$$\Delta v_{i,j} = \int_{t_i}^{t_j} \mathrm{Exp}(\bar{\omega}_{i,j}(t_j - t))dt\;\bar{\mathbf{a}}_{i,j} = \mathcal{E}_1(\bar{\omega}_{i,j}\Delta \mathrm{t}_{i,j})\;\bar{\mathbf{a}}_{i,j}\;\Delta \mathrm{t}_{i,j} \tag{9b}$$

$$\Delta p_{i,j} = \iint_{t_i}^{t_j} \mathrm{Exp}(\bar{\omega}_{i,j}(t_j - t))dt^2\;\bar{\mathbf{a}}_{i,j} = \mathcal{E}_2(\bar{\omega}_{i,j}\Delta \mathrm{t}_{i,j})\;\bar{\mathbf{a}}_{i,j}\;\Delta \mathrm{t}_{i,j}^2 \tag{9c}$$

with:

$$\mathcal{E}_1(\phi) = I_3 + \left(\frac{1\text{-}\cos(\|\phi\|)}{\|\phi\|^2}\right)\phi_\times + \left(\frac{\|\phi\|\text{-}\sin(\|\phi\|)}{\|\phi\|^3}\right)\phi_\times^2 \tag{9d}$$

$$\mathcal{E}_2(\phi) = \frac{1}{2}I_3 + \left(\frac{\|\phi\|\text{-}\sin(\|\phi\|)}{\|\phi\|^3}\right)\phi_\times + \left(\frac{1}{2\|\phi\|^2} + \frac{\cos(\|\phi\|)\text{-}1}{\|\phi\|^4}\right)\phi_\times^2 \tag{9e}$$

Using this model, the preintegration $\bar{\mathcal{I}}_{a,b}^{i,j}$ can thus be estimated by replacing $\Delta \mathrm{t}_{i,j}$ with $\Delta \mathrm{t}_{a,b}$ in equation 9:

$$\bar{\mathcal{I}}_{a,b}^{i,j} \doteq \begin{cases} \Delta\bar{R}_{a,b}^{i,j} = \mathrm{Exp}(\bar{\omega}_{i,j}\Delta \mathrm{t}_{a,b}) \\ \Delta\bar{v}_{a,b}^{i,j} = \mathcal{E}_1(\bar{\omega}_{i,j}\Delta \mathrm{t}_{a,b})\;\bar{\mathbf{a}}_{i,j}\;\Delta \mathrm{t}_{a,b} \\ \Delta\bar{p}_{a,b}^{i,j} = \mathcal{E}_2(\bar{\omega}_{i,j}\Delta \mathrm{t}_{a,b})\;\bar{\mathbf{a}}_{i,j}\;\Delta \mathrm{t}_{a,b}^2 \end{cases} \tag{10}$$

The constant values $\bar{\omega}_{i,j}$ and $\bar{\mathbf{a}}_{i,j}$ are deduced from $\mathcal{I}_{i,j}$ such that $\bar{\mathcal{I}}_{i,j}^{i,j} = \mathcal{I}_{i,j}$:

$$\bar{\omega}_{i,j} = \frac{\mathrm{Log}(\Delta R_{i,j})}{\Delta \mathrm{t}_{i,j}}, \tag{11a}$$

$$\bar{\mathbf{a}}_{i,j} = \frac{1}{\Delta \mathrm{t}_{i,j}}\left[\mathcal{E}_1(\bar{\omega}_{i,j}\Delta \mathrm{t}_{i,j})\right]^{-1}\Delta v_{i,j}, \tag{11b}$$

with:

$$\left[\mathcal{E}_1(\phi)\right]^{-1} = I_3 - \frac{1}{2}\phi_\times + \left(\frac{1}{\|\phi\|^2}\text{-}\frac{1+\cos(\|\phi\|)}{2\|\phi\|\sin(\|\phi\|)}\right)\phi_\times^2. \tag{11c}$$

3.2. TSIP measurement

Using the temporal interpolation of the previous section, we propose an approximation of preintegration measurement called TSIP that takes into account both time shift and IMU biases update.

Consider four consecutive time instants t_i, t_j, t_k & t_l and three preintegration measurements $\mathcal{I}_{i,j}$, $\mathcal{I}_{j,k}$ and $\mathcal{I}_{k,l}$. After computing approximations of these preintegration measurements (section 3.1), a time shift update $\delta\tau_s$ on $\mathcal{I}_{j,k}$ can be taken into account as follows (Fig. 1):

$$\hat{\mathcal{I}}_{j,k}(\delta\tau_s) \doteq \begin{cases} \bar{\mathcal{I}}_{j+\delta\tau_s,k}^{j,k} \oplus \bar{\mathcal{I}}_{k,k+\delta\tau_s}^{k,l} & \text{if } \delta\tau_s > 0 \\ \bar{\mathcal{I}}_{j+\delta\tau_s,j}^{i,j} \oplus \bar{\mathcal{I}}_{j,k+\delta\tau_s}^{j,k} & \text{if } \delta\tau_s < 0, \end{cases} \tag{12}$$

The addition $\oplus$ of two IMU preintegration measurement interpolations $\bar{\mathcal{I}}$ is deduced from the preintegration definition

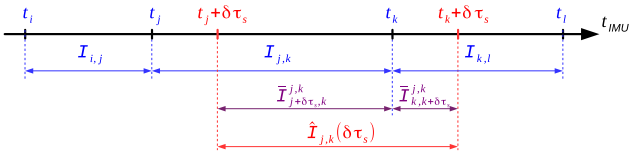

Figure 1: Time shifted preintegration measurement (equation 12) for a positive time shift update $\delta\tau_s$. The resulting TSIP measurement is a sum of two approximated IMU preintegration. The first one, $\bar{\mathcal{I}}^{j,k}_{j+\delta\tau_s,k}$, uses $\bar{\omega}_{j,k}$ and $\bar{\mathbf{a}}_{j,k}$ for a duration $\Delta\mathrm{t}_{j,k} - \delta\tau_s$, and the second one, $\bar{\mathcal{I}}^{k,l}_{k,k+\delta\tau_s}$, uses $\bar{\omega}_{k,l}$ and $\bar{\mathbf{a}}_{k,l}$ for a duration $\delta\tau_s$.

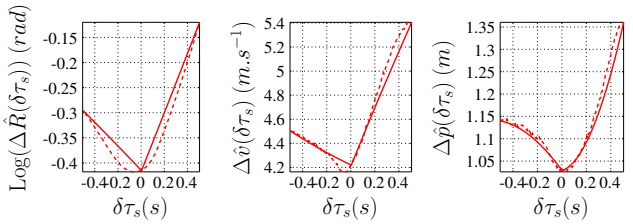

Figure 2: TSIP $\hat{\mathcal{I}}_{j,k}(\delta\tau_s)$ for $\delta\tau_s \in [-\Delta\mathrm{t}_{j,k}, \Delta\mathrm{t}_{j,k}]$ and $\Delta\mathrm{t}_{i,j} = \Delta\mathrm{t}_{j,k} = \Delta\mathrm{t}_{k,l} = 0.5s$ using IMU data from EuRoC MAV *V1_02_medium* sequence [2] (continuous line). Dotted lines are IMU preintegration values computed with raw IMU data. The first dimension of $\mathrm{Log}(\Delta\hat{R}_{j,k}(\delta\tau_s))$, $\Delta\hat{v}_{j,k}(\delta\tau_s)$ and $\Delta\hat{p}_{j,k}(\delta\tau_s)$ are here depicted. For the two other dimensions similar behavior is observed.

(equation 6) assuming that the biases are constant:

$$\bar{\mathcal{I}}^{i,j}_{a,b} \oplus \bar{\mathcal{I}}^{j,k}_{b,c} \doteq \begin{cases} \Delta\bar{R}^{i,j}_{a,b}\Delta\bar{R}^{j,k}_{b,c} \\ \Delta\bar{v}^{i,j}_{a,b} + \Delta\bar{R}^{i,j}_{a,b}\Delta\bar{v}^{j,k}_{b,c} \\ \Delta\bar{p}^{i,j}_{a,b} + \Delta\bar{v}^{i,j}_{a,b}\Delta\mathrm{t}_{b,c} + \Delta\bar{R}^{i,j}_{a,b}\Delta\bar{p}^{j,k}_{b,c} \end{cases} \tag{13}$$

A comparison of the TSIP $\hat{\mathcal{I}}_{j,k}(\delta\tau_s)$ and the preintegration recomputed using raw IMU data in $[t_j + \tau_s, t_k + \tau_s]$ is depicted in Figure 2. The TSIP measurement evolution with time shift influence sticks correctly to the reference value. In particular, the curves slightly bend in the right direction to stay close their real value.

Concerning IMU biases update, it is taken into account by substracting the update to gyroscope and accelerometer approximation for each interpolation of IMU preintegration $\bar{\mathcal{I}}$ defined in equation 10. Therefore, the following TSIP are an approximation of preintegration taking into account both time shift and IMU biases update without requiring to re-preintegrate gyroscope and accelerometer data:

$$\hat{\mathcal{I}}_{j,k}(\delta\tau_s, \delta\mathbf{b}_\mathrm{g}, \delta\mathbf{b}_\mathrm{a}) \doteq \tag{14}$$
$$\begin{cases} \bar{\mathcal{I}}^{j,k}_{j+\delta\tau_s,k}(\delta\mathbf{b}_\mathrm{g}, \delta\mathbf{b}_\mathrm{a}) \oplus \bar{\mathcal{I}}^{k,l}_{k,k+\delta\tau_s}(\delta\mathbf{b}_\mathrm{g}, \delta\mathbf{b}_\mathrm{a}) & \text{if } \delta\tau_s > 0, \\ \bar{\mathcal{I}}^{i,j}_{j+\delta\tau_s,j}(\delta\mathbf{b}_\mathrm{g}, \delta\mathbf{b}_\mathrm{a}) \oplus \bar{\mathcal{I}}^{j,k}_{j,k+\delta\tau_s}(\delta\mathbf{b}_\mathrm{g}, \delta\mathbf{b}_\mathrm{a}) & \text{if } \delta\tau_s < 0. \end{cases}$$

with every $\bar{\mathcal{I}}^{m,n}_{(.)}(\delta\mathbf{b}_\mathrm{g}, \delta\mathbf{b}_\mathrm{a})$ being $\bar{\mathcal{I}}^{m,n}_{(.)}$ of equation 10 where $\bar{\omega}_{m,n}$ and $\bar{\mathbf{a}}_{m,n}$ are replaced by $\bar{\omega}_{m,n} - \delta\mathbf{b}_\mathrm{g}$ and $\bar{\mathbf{a}}_{m,n} - \delta\mathbf{b}_\mathrm{a}$ respectively.

4. Incremental VISLAM initialization

Given the TSIP measurement presented in the previous section, we propose in this section an IMU initialization that aims to estimate both time shift and initial visual-inertial state in a robust non-linear optimization. After the presentation of previous work on VISLAM initialization, the proposed algorithm is explicited: optimization parameters are firstly analysed, then visual and inertial constraints constituting the VISLAM initialization graph are introduced, and finally its incremental optimization including the stopping criterion and the robustness to odometry failures is presented.

4.1. Related work

Since good knowledge of initial values of the system state has a significant impact on VISLAM performance, many methods to bootstrap visual-inertial odometry have already been proposed. In [12, 17], the use of known visual odometry and IMU data in a multi-step process enables to estimate IMU biases, gravity direction and metric scale of the monocular visual odometry. Recent work proposed by [3] extends a closed-form solution presented in [7, 4] to tightly fuse IMU data and visual features for a robust initialization procedure.These approaches estimate accuratly the inital state, however the time shift is not estimated. In [5] an IMU initialization is presented, that uses a constant angular velocity approximation on odometry to estimate time shift along with gyroscope bias ; the rest of the IMU initial state is then recovered separately in a second step.

Besides, all these initialization algorithms assume a continuity in the visual data. They do not handle failure situation of visual odometry (abrupt movement, motion blur). Furthermore, their criteria for stopping the IMU initialization do not guarantee the quality of the estimation. An initialization where odometry failures are handled and with a reliable end of estimation criteria is therefore presented.

4.2. Optimization parameters

Using visual odometry poses and TSIP measurement described in section 3, the proposed VISLAM initialization algorithm aims to give a first estimation of the constant time shift τ_s and the initial IMU state *i.e.* the IMU biases $\mathbf{b}_\mathrm{g}$ and $\mathbf{b}_\mathrm{a}$ and the gravity direction. The scale factor s of the visual odometry may also be estimated for monocular SLAM odometry. Since gravity vector norm is known, only gravity direction has to be estimated. Thus gravity vector can be written as ${}_\mathrm{w}\mathbf{g} = G\,\mathrm{Exp}(\psi)\mathbf{u}$, with G the gravity norm, $\mathbf{u}$ a unit vector and ψ the vector reflecting the gravity direction.

Besides, the rotation component of each odometry pose is added in the optimization. Indeed, since gravity is ex-

pressed in the world reference frame, a rotational drift of the odometry would lead to an inaccurate estimation of the gravity. Only the first rotation is fixed to remove gauge freedom. In contrast, translation components of the odometry poses remain fixed, since only delta positions are used (and their covariance) to estimate the initialization parameters. Velocity is not included in the optimization since VSLAM does not provide accurate velocity estimation. It also reduces the dimension of the estimated parameters. Consequently, the set of optimized parameters for $N+1$ odometry keyframes is defined as:

$$\theta \doteq \{\tau_s, \mathbf{b}_\mathrm{g}, \mathbf{b}_\mathrm{a}, s, {}_\mathrm{W}\mathbf{g}\} \cup \{R_i\}_{i=1\ldots N} \tag{15}$$

The extrinsic calibration of the IMU is considered fixed. It could be easily added in the proposed graph if required. The velocity ${}_\mathrm{W}\mathbf{v}_\mathrm{I}(t)$ of the IMU frame I expressed in frame W at any time t may also be computed at the end of the initialization by using equations 5, as done in [12].

4.3. Graph constraints

Considering $N+1$ odometry keyframes, the estimation of θ is obtained through the minimization of the sum of three different constraints:

$$\theta^* = \arg\min_\theta \left[\sum_{i=1}^{N} \mathbf{E}^{\text{bgts}}_{i,i+1} + \sum_{i=1}^{N-1} \mathbf{E}^{\text{sgbats}}_{i,i+1,i+2} + \sum_{i=1}^{N} \mathbf{E}^{\text{inter}}_{i,i+1} \right]. \tag{16}$$

where a constraint $\mathbf{E}^{(.)}$ corresponds to a residual error $\mathbf{r}^{(.)}$ weighted by its information matrix ${\Sigma^{(.)}}^{-1}$:

$$\mathbf{E}^{(.)} = {\mathbf{r}^{(.)}}^\top {\Sigma^{(.)}}^{-1} \mathbf{r}^{(.)}. \tag{17}$$

Here, $\mathbf{E}^{\text{bgts}}_{i,i+1}$ is a constraint between two odometry rotations and the rotation component of TSIP, enabling to refine both $\mathbf{b}_\mathrm{g}$ and τ_s. $\mathbf{E}^{\text{sgbats}}_{i,i+1,i+2}$ is the constraint connecting the scale factor s, the gravity vector ${}_\mathrm{W}\mathbf{g}$, the accelerometer bias $\mathbf{b}_\mathrm{a}$ and the time shift τ_s, using three consecutive odometry rotations of indices i, $i+1$, $i+2$ and their related TSIP. $\mathbf{E}^{\text{inter}}_{i,i+1}$ is the inter rotation prior term for two consecutive odometry rotations. Figure 3 illustrates the graph structure with the constraints and their related optimized variables.

The first two constraints are extensions of IMU initialization constraints used in [12] where TSIP method is applied. They are sufficient to estimate both the time shift and initial IMU parameters. The third constraint is added to the graph to increase the robusness of the proposed initialization to rotation drift of visual odometry.

4.3.1 BgTs constraint

BgTs residual error is obtained by incorporating TSIP measurement in equation 7a, and results in:

$$\mathbf{r}^{\text{bgts}}_{i,j} = \mathrm{Log}\left(\left[\Delta\hat{R}_{i,j}(\delta\tau_s, \delta\mathbf{b}_\mathrm{g}) \right]^\top R_\mathrm{BI}^\top R_j^\top R_i R_\mathrm{BI} \right). \tag{18}$$

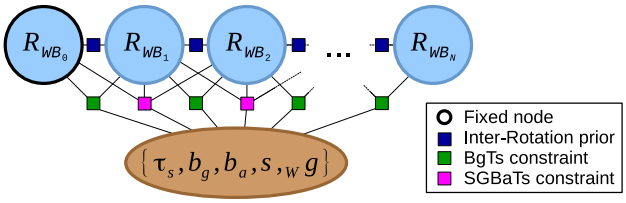

Figure 3: Structure of the graph used for the incremental initialization.

The information matrix of this residual is the inverse of the covariance of $\Delta R_{i,j}$, deduced from the covariance of $\mathcal{I}_{i,j}$:

$$\Sigma^{\text{bgts}}_{i,j} = [\mathbf{I}_{3\times3} \;\; \mathbf{0}_{3\times3} \;\; \mathbf{0}_{3\times3}] \; \Sigma^{\mathcal{I}_{i,j}} \; [\mathbf{I}_{3\times3} \;\; \mathbf{0}_{3\times3} \;\; \mathbf{0}_{3\times3}]^\top. \tag{19}$$

4.3.2 SGBaTs constraint

Residual error $\mathbf{r}^{\Delta p}$ (eq. 7c) is the relation between the scale factor s, the gravity vector and the preintegration measurements. However it involves state velocity ${}_\mathrm{W}\mathbf{v}_\mathrm{I}$, that does not belong to the set of parameters to optimize and that can not be fixed in the optimization since no precise estimate is known. Therefore, two residual errors $\mathbf{r}^{\Delta p}$ for three consecutives odometry keyframes at times t_i, t_j and t_k are combined to build the SGBaTs residual error:

$$\begin{aligned}
\mathbf{r}^{\text{sgbats}}_{i,j,k} &= \mathbf{r}^{\Delta p}_{j,k}\Delta\mathrm{t}_{i,j} - \mathbf{r}^{\Delta p}_{i,j}\Delta\mathrm{t}_{j,k} \\
&= s\left((\mathbf{p}_k - \mathbf{p}_j)\Delta\mathrm{t}_{i,j} - (\mathbf{p}_j - \mathbf{p}_i)\Delta\mathrm{t}_{j,k}\right) \\
&- ({}_\mathrm{W}\mathbf{v}_\mathrm{I}(t_j) - {}_\mathrm{W}\mathbf{v}_\mathrm{I}(t_i))\Delta\mathrm{t}_{i,j}\Delta\mathrm{t}_{j,k} \\
&+ \frac{1}{2}\,{}_\mathrm{W}\mathbf{g}(\Delta\mathrm{t}^2_{i,j}\Delta\mathrm{t}_{j,k} - \Delta\mathrm{t}^2_{j,k}\Delta\mathrm{t}_{i,j}) \\
&- R_j R_\mathrm{BI}\Delta p_{j,k}\Delta\mathrm{t}_{i,j} + R_i R_\mathrm{BI}\Delta p_{i,j}\Delta\mathrm{t}_{j,k} \\
&- (R_j - R_i)_\mathrm{B}\mathbf{p}_\mathrm{I}\Delta\mathrm{t}_{j,k} + (R_k - R_j)_\mathrm{B}\mathbf{p}_\mathrm{I}\Delta\mathrm{t}_{i,j}.
\end{aligned} \tag{20}$$

Since ${}_\mathrm{W}\mathbf{v}_\mathrm{I}(t_j) - {}_\mathrm{W}\mathbf{v}_\mathrm{I}(t_i)$ is equal to ${}_\mathrm{W}\mathbf{g}\,\Delta\mathrm{t}_{i,j} + R_i R_\mathrm{BI}\Delta v_{i,j}$ (eq. 5b), the presence of the velocity in equation 20 can be removed.

TSIP are substituted to preintegration measurement and SGBaTs residual becomes:

$$\begin{aligned}
\mathbf{r}^{\text{sgbats}}_{i,j,k} &= s\left((\mathbf{p}_k - \mathbf{p}_j)\Delta\mathrm{t}_{i,j} - (\mathbf{p}_j - \mathbf{p}_i)\Delta\mathrm{t}_{j,k}\right) \\
&- R_i R_\mathrm{BI}\Delta\hat{v}_{i,j}(\delta\tau_s, \delta\mathbf{b}_\mathrm{g}, \delta\mathbf{b}_\mathrm{a})\,\Delta\mathrm{t}_{i,j}\Delta\mathrm{t}_{j,k} \\
&- \frac{1}{2}\,{}_\mathrm{W}\mathbf{g}(\Delta\mathrm{t}^2_{i,j}\Delta\mathrm{t}_{j,k} + \Delta\mathrm{t}^2_{j,k}\Delta\mathrm{t}_{i,j}) \\
&+ R_i R_\mathrm{BI}\Delta\hat{p}_{i,j}(\delta\tau_s, \delta\mathbf{b}_\mathrm{g}, \delta\mathbf{b}_\mathrm{a})\,\Delta\mathrm{t}_{j,k} \\
&+ R_j R_\mathrm{BI}\Delta\hat{p}_{j,k}(\delta\tau_s, \delta\mathbf{b}_\mathrm{g}, \delta\mathbf{b}_\mathrm{a})\,\Delta\mathrm{t}_{i,j} \\
&- (R_j - R_i)_\mathrm{B}\mathbf{p}_\mathrm{I}\Delta\mathrm{t}_{j,k} + (R_k - R_j)_\mathrm{B}\mathbf{p}_\mathrm{I}\Delta\mathrm{t}_{i,j}.
\end{aligned} \tag{21}$$

To compute $\Sigma^{\text{sgbats}}_{i,j,k}$, we derive $\mathbf{r}^{\text{sgbats}}_{i,j,k}$ with respect to preintegration measurements and visual odometry delta positions. The covariance is then a sum of covariances from

visual and IMU measurements:

$$\Sigma^{\text{sgbats}}_{i,j,k} = J^{\mathcal{I}}_{i,j} \Sigma^{\mathcal{I}_{i,j}} {J^{\mathcal{I}}_{i,j}}^{\top} + J^{\mathcal{I}}_{j,k} \Sigma^{\mathcal{I}_{j,k}} {J^{\mathcal{I}}_{j,k}}^{\top} + J^{\mathcal{P}}_{i,j} \Sigma^{\Delta P_{i,j}} {J^{\mathcal{P}}_{i,j}}^{\top} + J^{\mathcal{P}}_{j,k} \Sigma^{\Delta P_{j,k}} {J^{\mathcal{P}}_{j,k}}^{\top}, \quad (22)$$

where jacobians $J^{\mathcal{I}}_{i,j}$, $J^{\mathcal{I}}_{j,k}$, $J^{\mathcal{P}}_{i,j}$ and $J^{\mathcal{P}}_{j,k}$ are analytically computed.

4.3.3 Inter rotation constraint

The uncertainty of the visual odometry rotations is included in the graph through the inter rotation covariance. Considering two consectutive rotations at indices i and j, these constraints act like a prior on their inter rotation $R_j^\top R_i$ related to its initial value $\tilde{R}_j^\top \tilde{R}_i$ determined by the odometry. The residual error is therefore the difference of these two rotations in Lie group $SO(3)$:

$$\mathbf{r}^{\text{inter}}_{i,j} = \text{Log}\left([\tilde{R}_j^\top \tilde{R}_i]^\top R_j^\top R_i \right). \quad (23)$$

The information matrix of this residual error is the inverse of the rotation component of the computed inter rotation covariance presented in equation 2:

$$\Sigma^{\text{inter}}_{i,j} = \begin{bmatrix} \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} \end{bmatrix} \Sigma^{\Delta P_{i,j}} \begin{bmatrix} \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} \end{bmatrix}^\top. \quad (24)$$

4.4. Incremental graph optimization

In this section, we describe how the graph presented above is initialized, incrementally constructed and optimized.

Optimization of the graph begin when the system is observable. For N odometry poses, since $\dim(\theta) = 3N + 7$ and $\dim(\mathbf{r}^{\text{inter}}_{i,j}) + \dim(\mathbf{r}^{\text{bgts}}_{i,j}) + \dim(\mathbf{r}^{\text{sgbats}}_{i,j,k}) = 9N - 12$, the minimum number of odometry poses needed before beginning estimation is equal to 4. The graph is optimized using Levenberg-Marquardt as non-linear solver. The optimization then takes place as follow:

- **Initialization and first optimization of the graph:** The first odometry rotations are added to the graph and IMU biases and time shift are set to zero. IMU data are preintegrated between timestamps of consecutive odometry key-frames poses and TSIP measurements computed (equation 11). A first estimation of $\mathbf{b}_\text{g}$ and τ_s is obtained by optimizing the graph with BgTs and delta rotation prior constraints. The scale and gravity vector are then roughly evaluated through the resolution of a linear system as equation 12 in [12]. The SGBaTs constraints are finally added to the graph to refine all θ in a single optimization.

- **Incremental optimization:** When a new odometry pose and its associated IMU data are available, the graph is incrementally completed. First, new IMU data are preintegrated and the TSIP between the previous and the new odometry pose is estimated. Then constraints related to new inputs are added to the graph: inter rotation, BgTs and SGBaTs, and a new graph optimization is launched to refine θ.

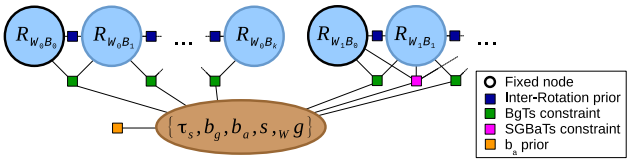

Figure 4: Graph structure after an odometry failure at time instant t_k. Here W_0 is the fomer world coordinate frame before odometry failure and W_1 is the new one. Former BgTs edges are still valid and are kept in the optimization. However former SGBaTs edges have to be removed and a prior on $\mathbf{b}_\text{a}$ is added to the optimization.

4.5. Robustness to odometry failure

In case of an odometry failure, a new arbitrary world frame is set by the odometry (in case of no relocalisation techniques). Since not all components of θ become obsolete, a complete reset of the initialization would lead to a loss of information.

In the optimization graph, the former scale factor and gravity vector direction estimated become obsolete as well as all SGBaTs constraints. BgTs constraints are independent of the choice of world coordinate frame and thus remain valid. Therefore, gyrometer bias and time shift estimations are not affected by odometry failure. Concerning accelerometer bias, since SGBaTs constraints have to be removed from the graph, its former estimation is resumed by marginalization and this information matrix is used as a prior to next estimations.

The optimization continues using the new visual odometry poses, by fixing the first rotation to remove gauge freedom, as depicted in Figure 4.

4.6. Stopping criterion

After each graph optimization, the uncertainty $\Sigma^{\mathcal{Z}}$ of initialization parameters set $\mathcal{Z} = \theta \setminus \{R_i\}_{i=1\dots N}$ is obtained by marginalization. A standard deviation σ is assigned to every component of $\mathcal{Z}$, representing the accuracy level desired for this component. This enables to normalize $\Sigma^{\mathcal{Z}}$ into a unitless covariance $\Sigma^{stop} = M_\sigma^{-1}\, \Sigma^{\mathcal{Z}}\, M_\sigma^{-1}$ with $M_\sigma = diag(\sigma_{\tau_s}, \sigma_{\mathbf{b}_\text{g}}, \sigma_{\mathbf{b}_\text{a}}, \sigma_s, \sigma_{{}_W\mathbf{g}})$. Therefore, all components in $\mathcal{Z}$ are accurate enough when the biggest eigen value of Σ^{stop} is inferior to 1.

We use $\sigma_{\tau_s} = 1ms$, $\sigma_{\mathbf{b}_\text{g}} = 0.0005 rad.s^{-1}$, $\sigma_{\mathbf{b}_\text{a}} = 0.01 m.s^{-2}$, $\sigma_s = 0.01$ and $\sigma_{{}_W\mathbf{g}_{dir}} = 0.01 rad$ in experiments. Note that these values are too optimistic: this over-

confidence is probably due to $\Sigma^{\mathcal{Z}}$ being inconsistent. Indeed recent work on invariant filters [1] indicates that any classical SLAM has an overconfident state estimate, so the VSLAM we used as input is no exception to this statement.

5. Experiments

In this section, the accuracy and the robustness of the proposed initialization is demonstrated.

First, the accuracy of the temporal calibration and the initial IMU state estimated with the proposed initialization is studied. Then, its impact on VISLAM peformance is presented and compared with the initialization described in [12]. Finaly, the robustness of the proposed initialization to odometry failures is demonstrated.

We use g2o [8] to create and optimize the graph, with custom edges and vertices implementations. Experiments are perfomed on Visual-Inertial Datasets EuRoC MAV [2] and TUM VI Benchmark [15]. The VISLAM algorithm used to evaluate the proposed initialization is a non-linear optimization based VISLAM similar to the one described in [6].

5.1. Accuracy of time shift and IMU state estimation

Temporal calibration based on TSIP measurements is evaluated by simulating time offsets on IMU data of the *V1_02_medium* sequence. About fourty initializations, distributed on the whole sequence, have been performed for every simulated time shift in range from -500ms to 500ms. Results of time shift estimation are depicted in Figure 5. The error of the estimated time shift is constantly under 1% of the simulated value and less than 2ms in range [-100ms, 100ms]. Average error is under 0.5ms regardless of the time shift value. We thus have a better temporal calibration than the initialization proposed in [5], where estimations errors lay between -3ms and 3ms for this sequence. The initialization duration is below 9 seconds on average and does not depend on the time shift value.

To demonstrate that IMU state and time shift have to be estimated simultaneously, we compare in Figure 6 the gyrometer bias estimation obtained by the initialization method proposed in [12] that does not estimate time shift with the one obtained by the proposed initialization. $\mathbf{b}_g$ error is nearly linear with time shift, while with the proposed initialization the error has an almost constant value around 2% of the $\mathbf{b}_g$ norm. For the other IMU state parameters we observe a similar behavior. This illustrates that IMU initialization without temporal calibration fails to provide correct estimation of the inital IMU state when time shift occurs.

5.2. Initialization impact on VISLAM performance

The performance of the proposed VISLAM initialization is demonstrated on different public dataset. Results are resumed in Table 1. Time shifts of 10, 20 and 30ms are simulated, and the proposed initialization is compared to [12] by measuring the VISLAM final position error using dataset ground truth: $err = \|\mathbf{p}_N - \mathbf{p}_0 - R_0^\top R_N {R_0^{\text{GT}}}^\top (\mathbf{p}_N^{\text{GT}} - \mathbf{p}_0^{\text{GT}})\|$, with $\mathbf{p}_0$ and $\mathbf{p}_N$ being the first and last positions of the VISLAM. The VISLAM localization accuracy obtained with the proposed initialization is stable regardless of the time shift value, whereas when using initialization of [12], it deteriorates with increasing time shift. On the four sequences of Table 1, VISLAM average error is 1,05m using the proposed initialization and 2.87m with the initialization of [12].

Our own sequence acquired with two camera FLIR BlackFly S and an 6-axis SBG Ellipse-N IMU is also used to evaluate the VISLAM accuracy with the proposed initialization. Images of this sequence is illustrated in Figure 7. The trajectory forms a rectangle of 25 meters wide and 55 meters long in the corridors of a building. The starting and the ending positions are the same. The VISLAM localization accuracy is measured as the distance between the last

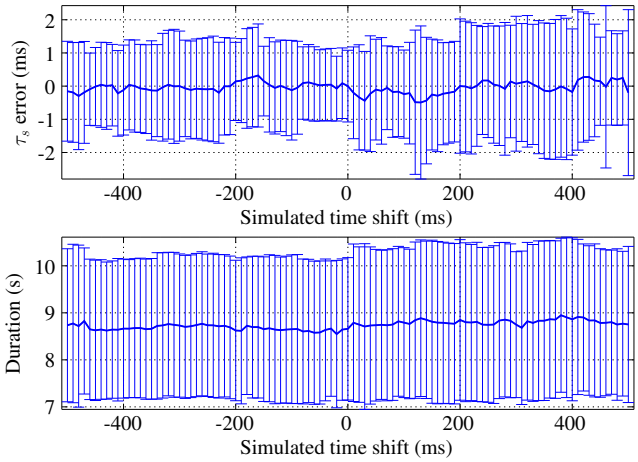

Figure 5: Mean and standard deviation of time shift error and duration of the proposed VISLAM initialization for each simulated time shift, on EuRoC MAV *V1_02_medium* sequence [2].

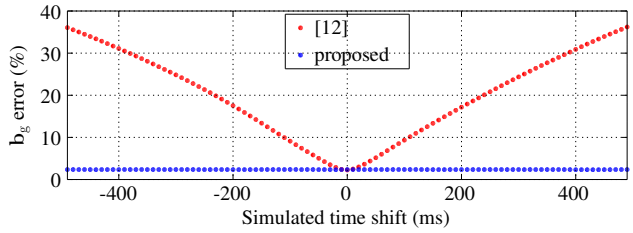

Figure 6: Mean of gyrometer bias error for each simulated time shift on the EuRoC MAV *V1_02_medium* sequence [2] estimated by the VISLAM initialization of [12] (red) and by the one we propose (blue).

Table 1: Evaluation of the VISLAM accuracy according to the initialization algorithm for different simulated time shift values

Sequence	Time shift (ms)	Final position error (m) Proposed	[12]
TUM Slides1	10	**1.7190**	4.5410
	20	**2.3075**	7.7297
	30	**2.0917**	5.3901
TUM Corridor1	10	**1.3026**	2.1867
	20	**1.3162**	3.5366
	30	**1.3071**	5.8845
EuRoC MH3	10	0.32065	**0.11650**
	20	**0.33221**	0.70788
	30	**0.33981**	2.2444
EuRoC MH5	10	**0.53077**	0.57704
	20	**0.53914**	0.53983
	30	**0.54026**	0.96565

estimated position and the first one. At the beginning of the sequence, an april target is filmed to obtain a ground truth of the time shift with Kalibr [14]. The time shift estimated by Kalibr is 22.36ms whereas its estimation by our initialization is 21.97ms. With our initialization, the VISLAM localization error is 0.23% of the trajectory length while with the initialization of [12] it rises to 1.5%. The VISLAM localization that uses the proposed initialization is illlustrated in Figure 7.

5.3. Initialization robustness

We here demonstrate the robustness of the proposed VISLAM initialization to odometry failures on the *V1_03_difficult* sequence with a simulated time shift of 50ms. On this sequence, the monocular VSLAM algorithm crashes after 5 seconds and restarts seconds later. The proposed initialization with and without handling odometry failures is launched. Gyrometer bias, gravity direction and time shift estimation have converged before the VSLAM crash ($\tau_s = 49.91ms$, $\mathbf{b}_\mathrm{g}$ error $= 1.56\%$ of the $\mathbf{b}_\mathrm{g}$ norm), only the accelerometer bias estimation did not have enough time to converge. Figure 8 shows the evolution of the accelerometer bias error during both initialization. With odometry failures handling, $\mathbf{b}_\mathrm{a}$ converges after 10 seconds of the sequence whereas by restarting the initialization after the VSLAM crash it takes more than 14 seconds. Adding a prior on $\mathbf{b}_\mathrm{a}$ in the graph (Figure 4) allows continuity in the VISLAM initialization if VSLAM failure occurs.

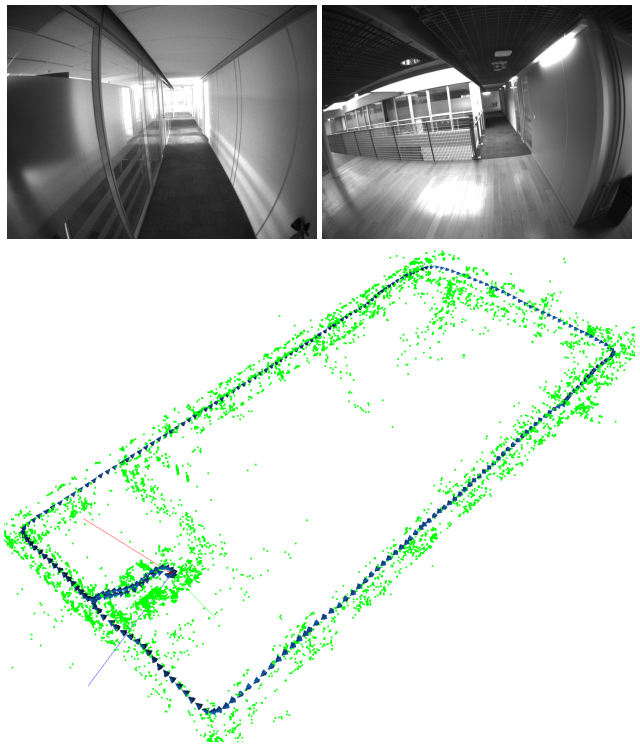

Figure 7: Top: illustration of the sequence we acquired in our building. Bottom : VISLAM localization and reconstuction that has been initialized with the proposed algorithm described in Section 4.

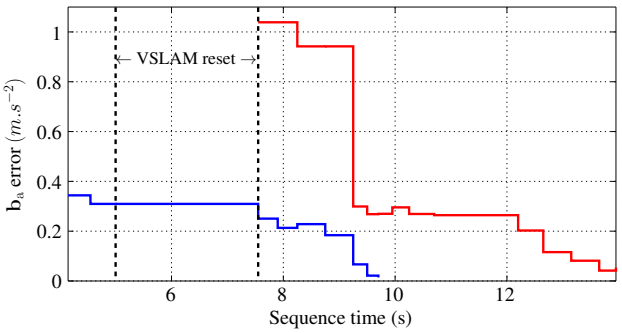

Figure 8: Evolution of the accelerometer bias error during the proposed incremental initialization with (blue) and without (red) handling odometry failures on the EuRoC MAV *V1_03_difficult* sequence [2]. Without odometry failures handling the VISLAM initialization is restarted.

6. Conclusion

In this paper is presented an online incremental VISLAM initialization that estimates both time shift and initial IMU state in a single non-linear optimization. To this end we introduced TSIP measurements: a new concept for temporal interpolation of IMU preintegration. Experiments on real data have demonstrated the precision and the robustness of the proposed initialization algorithm thus enabling to improve VISLAM accuracy.

The TSIP measurement is a generic framework that could be used in closed form initialization [3] or for any sensors system including an IMU (*e.g.* IMU/LIDAR).

References

[1] A. Barrau and S. Bonnabel. An EKF-SLAM algorithm with consistency properties. *CoRR*, abs/1510.06263, 10 2015. 7

[2] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. Achtelik, and R. Siegwart. The EuRoC micro aerial vehicle datasets. *The International Journal of Robotics Research*, 35(10):1157–1163, 2016. 4, 7, 8

[3] C. Campos, J. M. M. Montiel, and J. D. Tards. Fast and robust initialization for visual-inertial SLAM. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 1288–1294, 2019. 4, 8

[4] J. Domínguez-Conti, J. Yin, Y. Alami, and J. Civera. Visual-inertial slam initialization: A general linear formulation and a gravity-observing non-linear optimization. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 37–45. IEEE, 2018. 4

[5] Z. Feng, J. Li, Z. Lundong, and C. Chen. Online spatial and temporal calibration for monocular direct visual-inertial odometry. *Sensors*, 19:2273, 05 2019. 1, 4, 7

[6] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza. IMU preintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation. Georgia Institute of Technology, 2015. 1, 2, 3, 7

[7] J. Kaiser, A. Martinelli, F. Fontana, and D. Scaramuzza. Simultaneous state initialization and gyroscope bias calibration in visual inertial aided navigation. *IEEE Robotics and Automation Letters*, 2(1):18–25, 2016. 4

[8] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard. g2o: A general framework for graph optimization. In *2011 IEEE International Conference on Robotics and Automation*, pages 3607–3613. IEEE, 2011. 7

[9] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale. Keyframe-based visual–inertial odometry using nonlinear optimization. *The International Journal of Robotics Research*, 34(3):314–334, 2015. 2

[10] A. I. Mourikis and S. I. Roumeliotis. A multi-state constraint kalman filter for vision-aided inertial navigation. In *Proceedings 2007 IEEE International Conference on Robotics and Automation*, pages 3565–3572. IEEE, 2007. 1

[11] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE transactions on robotics*, 31(5):1147–1163, 2015. 2

[12] R. Mur-Artal and J. D. Tards. Visual-inertial monocular SLAM with map reuse. *IEEE Robotics and Automation Letters*, 2(2):796–803, 2017. 1, 4, 5, 6, 7, 8

[13] T. Qin and S. Shen. Online temporal calibration for monocular visual-inertial systems. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3662–3669. IEEE, 2018. 1

[14] J. Rehder, J. Nikolic, T. Schneider, T. Hinzmann, and R. Siegwart. Extending kalibr: Calibrating the extrinsics of multiple IMUs and of individual axes. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4304–4311. IEEE, 2016. 1, 8

[15] D. Schubert, T. Goll, N. Demmel, V. Usenko, J. Stückler, and D. Cremers. The TUM VI benchmark for evaluating visual-inertial odometry. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1680–1687. IEEE, 2018. 7

[16] K. Sun, K. Mohta, B. Pfrommer, M. Watterson, S. Liu, Y. Mulgaonkar, C. J. Taylor, and V. Kumar. Robust stereo visual inertial odometry for fast autonomous flight. *IEEE Robotics and Automation Letters*, 3(2):965–972, 2018. 1

[17] P. L. Tong Qin and S. Shen. Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 2017. 4

[18] Y. Yang, P. Geneva, K. Eckenhoff, and G. Huang. Degenerate motion analysis for aided ins with online spatial and temporal sensor calibration. *IEEE Robotics and Automation Letters*, 4(2):2070–2077, 2019. 1