Good Features to Track for Visual SLAM

Guangcong Zhang
School of ECE, Georgia Tech.
zhanggc@gatech.edu

Patricio A. Vela
School of ECE, Georgia Tech.
pvela@gatech.edu

Abstract

Not all measured features in SLAM/SfM contribute to accurate localization during the estimation process, thus it is sensible to utilize only those that do. This paper describes a method for selecting a subset of features that are of high utility for localization in the SLAM/SfM estimation process. It is derived by examining the observability of SLAM and, being complimentary to the estimation process, it easily integrates into existing SLAM systems. The measure of estimation utility is formulated with temporal and instantaneous observability indices. Efficient computation strategies for the observability indices are described based on incremental singular value decomposition (SVD) and greedy selection for the temporal and instantaneous observability indices, respectively. The greedy selection is near-optimal since the observability index is (approximately) submodular. The proposed method improves localization and data association. Controlled synthetic experiments with ground truth demonstrate the improved localization accuracy, and real-time SLAM experiments demonstrate the improved data association.

1. Introduction

A fact in visual SLAM/SfM is that not all of the features being tracked contribute to accurate estimation of the camera poses and the map. Finding the features that provide the best values for estimation is important when SLAM is to be used for practical purposes. It is equally important when considering visual SLAM systems developed for large-scale/dense reconstruction with massive data processing needs [27, 8, 12, 19, 29].

Conventionally, a fully data-driven and randomized process like RANSAC is used to select the valuable features by retrieving the inlier set [11]. Various approaches [33, 7] were later proposed to improve the computational efficiency and robustness of RANSAC in visual SLAM. These methods are data-driven and make no use of structural information of the relative motion. Recent research efforts have sought more systematic criterion for selecting the valuable features. [5] propose to exploit the co-visibility of features by cameras to select the best subset of points, but this method is developed for Bundle Adjustment and requires the complete structure of features-camera graph as a priori knowledge. For SLAM, information gain has been a popular criterion for such a selection [9, 6, 17, 20]. The rationale behind information gain is that selecting the features which maximize the information gain in estimation will maximize the uncertainty reduction for both the camera pose and landmark positions. Nevertheless, low uncertainty in estimation is not equivalent to high accuracy. For instance, if drift exists in the estimate, the converged estimates with lowest uncertainty still suffer from the drift. Rather, the accuracy of the converged SLAM estimate is determined by the operator mapping the projective space of image observations to the space of camera motion and feature 3D positions, and its temporal dynamics, as indicated in the right block in Fig. 2. Intuition then indicates that the better conditioned this operator is, the more tolerant the output space is to the perturbations in the input space. This operator encodes the camera motion across frames due to temporal coupling of SLAM estimates.

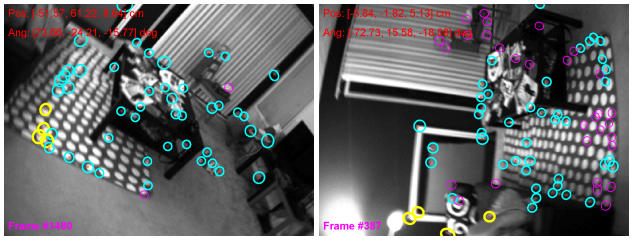

Figure 1. The proposed method selects the measurements (highlighted in yellow) which provide the most value to the SLAM estimation, by considering observability scores. In the example the camera is mostly rotating w.r.t. the optical axis.

To exploit nature of $SE\langle 3\rangle$ SLAM operator to feature ranking, we study the SLAM problem using system theory to define the **observability scores** for feature selection. System theory, especially observability theory, have been seen in robotics literature, but mostly restricted to 1D SLAM [14] or 2D(planar motion) SLAM [1, 24, 28] rather than monocular camera SLAM on $SE\langle 3\rangle$. Moreover, ob-

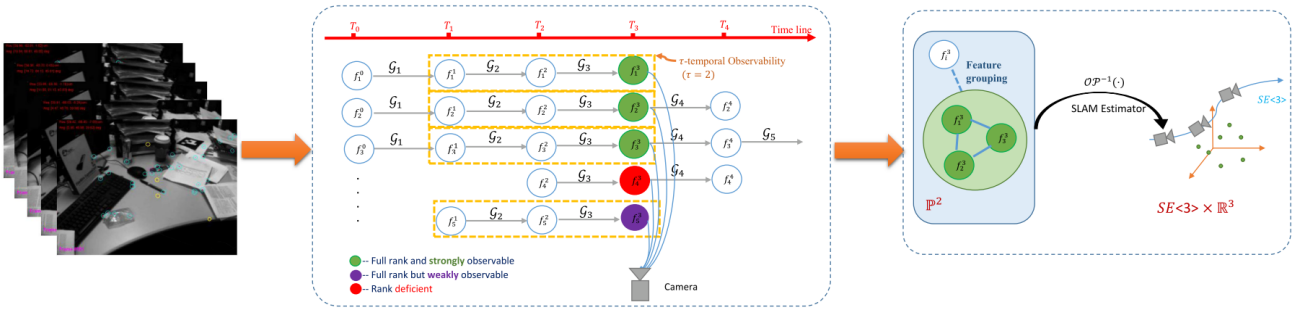

Figure 2. Overview of our approach. The proposed method can be plugged in as a sub-step in the SLAM process. In a time step (T_3 in the figure), for features which are initially matched, we first examine the rank conditions for them, i.e. whether the feature is **completely observable** to the SLAM system. If rank condition a feature is satisfied (depicted in green/purple), the τ-temporal observability score is evaluated by considering the relative motion of the feature in the past τ local frames. Features with high observability scores are selected as 'good' features (depicted in green). If the number of highly observable features is too few, feature grouping with a submodular learning scheme is applied to collect more good features. These subset of good features provide the near-optimal value for SLAM estimation.

servability theory has mainly been for full rank observability condition analysis, much as in [34] which analyzes bearings-only SLAM. For visual SLAM on $SE\langle 3\rangle$, [26] provides an analysis of observability, but it is for stereo-vision SLAM with planar displacement. [32] discusses observability tests for camera ego-motion from perspective views at time instances. Few works use observability in algorithm design rather than merely observable condition analysis /observability tests. [18] presented a framework for improving the consistency of EKF-based *planar SLAM* by finding linearization points that ensure the observable subspace is of appropriate dimension for the linearized system.

Contribution. Using systems theory, we develop a feature ranking criterion for selecting the features which provide good conditioning for visual SLAM ego-motion estimation. The overview of our method is depicted in Fig. 2. This paper has three major contributions: we (1) propose a feature ranking criterion based on **observability scores** using a **complete observability** condition for $SE\langle 3\rangle$ SLAM; (2) describe an efficient algorithm for computing temporal observability based on incremental SVD; and (3) describe an efficient algorithm for computing instantaneous observability via submodular learning. The algorithm is called the Good Features algorithm and can be integrated into most existing SLAM implementations to arrive at GF-SLAM. The contributions lead to performance gains regarding ego-motion estimation and data-association in visual SLAM, which are shown via experiments.

2. Good Features to Track for Visual SLAM

Let $\mathcal{F}$ be the set of features being tracked during the monocular SLAM process. Much like [30] sought good features within an image for data association across frames, the *Good Features* algorithm here aims to find the subset of features which aids most the SLAM camera ego-motion estimates across time (in terms of accuracy and robustness to noise). This subset is selected by ranking features according to their contribution to system observability (higher system observability means better conditioned estimation). In order to formulate the ranking score for each feature, the SLAM system is first modeled with $SE\langle 3\rangle$ motion. The score is then formulated based on the observability of the subsystem composed of the camera and each individual feature.

2.1. Motion and Observations for $SE\langle 3\rangle$ SLAM

Here, the SLAM scenario with features and anchors is considered. The SLAM system dynamics are modeled under the hybrid $SE\langle 3\rangle$ state common to robotics (position in world frame $\mathcal{W}$, with orientation in body frame $\mathcal{R}$) [25], with a perspective camera measurement model.

2.1.1 Dynamic and Measurement Models

For a system with discrete observations, a constant velocity motion model suffices [11]. Accordingly, given the $SE\langle 3\rangle$ position and orientation $\mathbf{r}^{\mathcal{W}}_{\mathcal{R}_k}$, $\mathbf{q}^{\mathcal{W}}_{\mathcal{R}_k}$ (vector, quaternion), and associated velocities $\mathbf{v}^{\mathcal{W}}_{\mathcal{R}_k}$, $\omega^{\mathcal{R}}$, at time k, the camera state

$$\mathbf{x}^{\mathcal{W}}_{\mathcal{R}_k} = \begin{pmatrix} \mathbf{r}^{\mathcal{W}}_{\mathcal{R}_k} & \mathbf{q}^{\mathcal{W}}_{\mathcal{R}_k} & \mathbf{v}^{\mathcal{W}}_{\mathcal{R}_k} & \omega^{\mathcal{R}} \end{pmatrix}^{\top}$$

is updated per:

$$\mathbf{x}^{\mathcal{W}}_{\mathcal{R}_{k+1}} = \begin{pmatrix} \mathbf{r}^{\mathcal{W}}_{\mathcal{R}_k} + (\mathbf{v}^{\mathcal{W}}_{\mathcal{R}_k} + \mathbf{V}^{\mathcal{W}})\Delta t \\ \mathbf{q}^{\mathcal{W}}_{\mathcal{R}_k} \times \exp\left(\left[\omega^{\mathcal{R}} + \mathbf{\Omega}^{\mathcal{R}}\right]\Delta t\right) \\ \mathbf{v}^{\mathcal{W}}_{\mathcal{R}_k} + \mathbf{V}^{\mathcal{W}} \\ \omega^{\mathcal{R}} + \mathbf{\Omega}^{\mathcal{R}} \end{pmatrix},$$

where $\mathbf{V}^{\mathcal{W}}, \mathbf{\Omega}^{\mathcal{R}}$ are zero-mean Gaussian noise. The measurement model for the i-th feature ${}^{(i)}\mathbf{p}^{\mathcal{W}}_k \in \mathbb{R}^3$ is pinhole

projection:

$$
{}^{(i)}\mathbf{p}^{\mathcal{R}_k} = [p_x^{\mathcal{R}_k}, p_y^{\mathcal{R}_k}, p_z^{\mathcal{R}_k}]^\top = \mathrm{R}_{\mathcal{W}}^{\mathcal{R}_k}\left(\left(\mathbf{q}_{\mathcal{R}_k}^{\mathcal{W}}\right)^{-1}\right)\left({}^{(i)}\mathbf{p}_k^{\mathcal{W}} - \mathbf{r}_{\mathcal{R}_k}^{\mathcal{W}}\right),
$$
$$
\mathbf{h}_i^{\mathcal{R}_k} = \mathrm{Distort}\left[\begin{pmatrix} u_0 - fk_u \frac{p_x^{\mathcal{R}_k}}{p_z^{\mathcal{R}_k}} \\ v_0 - fk_v \frac{fp_y}{p_z^{\mathcal{R}_k}} \end{pmatrix}\right], \tag{1}
$$

where $\mathrm{R}(\mathbf{q})$ is the rotation matrix of $\mathbf{q}$; fk_u, fk_v, u_0, v_0 are the camera intrinsic parameters; and $\mathrm{Distort}[\cdot]$ is nonlinear image distortion [10].

2.1.2 Piece-wise Linear System (PWLS) modeled for SLAM

Assume the system has N_f features and N_a anchors. An anchor is a 3D point in $\mathcal{W}$ whose position is known and is not included in estimation process, while a feature is 3D point whose position is not certain (at least initially). Both are observed by the camera as per Equation (1). For the k-th time segment $\mathcal{T}_k \equiv [t_k, t_{k+1})$ (from time k to time $k+1$), the dynamics of the whole system with input $\mathbf{u}_k$ are

$$
\mathbf{X}_{k+1}^{\mathcal{W}} \triangleq \begin{pmatrix} \mathbf{x}_{\mathcal{R}_{k+1}}^{\mathcal{W}} \\ \mathbf{P}_{k+1}^{\mathcal{W}} \end{pmatrix} = \mathrm{f}\left(\begin{pmatrix} \mathbf{x}_{\mathcal{R}_k}^{\mathcal{W}} \\ \mathbf{P}_k^{\mathcal{W}} \end{pmatrix} \middle| \mathbf{A}_k^{\mathcal{W}}\right) + \mathbf{u}_k, \tag{2a}
$$
$$
\mathbf{h}^{\mathcal{R}_{k+1}} = \mathrm{h}^{\mathcal{R}_{k+1}}\left(\begin{pmatrix} \mathbf{x}_{\mathcal{R}_k}^{\mathcal{W}} \\ \mathbf{P}_k^{\mathcal{W}} \end{pmatrix} \middle| \mathbf{A}_k^{\mathcal{W}}\right), \tag{2b}
$$

where $\mathbf{P}_k^{\mathcal{W}} \triangleq ({}^{(1)}\mathbf{p}_k^{\mathcal{W}}, {}^{(2)}\mathbf{p}_k^{\mathcal{W}}, ..., {}^{(N_f)}\mathbf{p}_k^{\mathcal{W}})^\top \in \mathbb{R}^{3N_f}$ is the map state vector by stacking the feature vectors, $\mathbf{A}_k^{\mathcal{W}} \triangleq ({}^{(1)}\mathbf{a}_k^{\mathcal{W}}, {}^{(2)}\mathbf{a}_k^{\mathcal{W}}, ..., {}^{(N_a)}\mathbf{a}_k^{\mathcal{W}})^\top \in \mathbb{R}^{3N_a}$ is the anchor state vector, and $\mathbf{h}^{\mathcal{R}_{k+1}} \triangleq (\mathbf{h}_1^{\mathcal{R}_{k+1}}, \mathbf{h}_2^{\mathcal{R}_{k+1}}, ..., \mathbf{h}_N^{\mathcal{R}_{k+1}})^\top \in \mathbb{I}^{2(N_f+N_a)}$ is the measurement vector at time $k+1$ with measurements from both features and anchors.

With the smooth motion assumption, the system at $\mathcal{T}_k$ is linearized via $\mathbf{X}_{k+1}^{\mathcal{W}} \approxeq \mathbf{X}_k^{\mathcal{W}} + \mathrm{D}f\left(\mathbf{X}_k^{\mathcal{W}}\right) \cdot \mathbf{X}_k^{\mathcal{W}} + \mathbf{u}_k$, $\mathbf{h}^{\mathcal{R}_{k+1}} \approxeq \mathbf{h}^{\mathcal{R}_k} + \mathrm{D}\mathrm{h}^{\mathcal{R}_{k+1}}\left(\mathbf{X}_k^{\mathcal{W}}\right) \cdot \mathbf{X}_k^{\mathcal{W}}$. The linearized systems across time segments form the **piece-wise linear system (PWLS)** in (3) below, which approximates the time-varying system in (2).

$$
\begin{cases} \mathbf{X}_{k+1}^{\mathcal{W}} = \mathrm{F}^{\mathcal{R}_k}\mathbf{X}_k^{\mathcal{W}} + \mathbf{u}_k \\ \delta\mathbf{h}^{\mathcal{R}_k} = \mathrm{H}^{\mathcal{R}_k}\mathbf{X}_k^{\mathcal{W}} \end{cases} \quad \text{for } t \in \mathcal{T}_k \tag{3}
$$

The PWLS preserves the characteristic behavior of the original time-varying system with little loss of accuracy [15].

2.2. System Observability Measure

For a discrete PWLS, the sufficient and necessary condition for the system to be completely observable is given by the following Lemma:

Lamma 1. *[15] A discrete PWLS is completely observable iff the Total Observability Matrix (TOM) is full-rank.*

$$
\mathcal{Q}_{\mathrm{TOM}}(j) = \begin{pmatrix} \mathcal{Q}_1 \\ \mathcal{Q}_2 F_1^{n-1} \\ \vdots \\ \mathcal{Q}_r F_{r-1}^{n-1} F_{r-2}^{n-1} \cdots F_1^{n-1} \end{pmatrix} \tag{4}
$$

where F_j is the process matrix and H_j is the measurement matrix for time segment j. $\mathcal{Q}_j$ is the linear observability matrix, $\mathcal{Q}_j^\top = \left[H_j^\top | (H_j F_j)^\top | \cdots | (H_j F_j^{n-1})^\top\right]$.

Computation of the TOM is expensive. However, for the SLAM system described in Equation 3, $\mathcal{N}(\mathcal{Q}_j) \subset \mathcal{N}(F_j)$. Lemma 2 provides a proxy to examine the full rank condition of the system

Lamma 2. *[15] For PWLS, when $\mathcal{N}(\mathcal{Q}_j) \subset \mathcal{N}(F_j)$, the stripped Observability Matrix (SOM)*

$$
\mathcal{Q}_{\mathrm{SOM}}(j) = \left[\mathcal{Q}_1^\top | \mathcal{Q}_2^\top | \cdots | \mathcal{Q}_j^\top\right]^\top. \tag{5}
$$

has the same nullspace as TOM, i.e. $\mathcal{N}(\mathcal{Q}_{\mathrm{SOM}}(j)) = \mathcal{N}(\mathcal{Q}_{\mathrm{TOM}}(j))$.

Theorem 1. *When $N_f = 0$, a necessary condition for system (3) to be completely observable within J is (1) $J = 1$ and $N_a \geq 3$, or (2) $J \geq 2$ and $N_a \geq 1$.*

Proof. The SLAM system with N_f features and N_a anchors has the PWLS matrices

$$
\mathrm{F}^{\mathcal{R}_k} = \begin{pmatrix} \mathrm{F}_{\mathbf{x}_{\mathcal{R}_k}^{\mathcal{W}}} & \mathbf{0}_{13\times 3N_f} \\ \mathbf{0}_{3N_f\times 13} & \mathbf{I}_{3N_f\times 3N_f} \end{pmatrix}, \tag{6}
$$

and

$$
\mathrm{F}_{\mathbf{x}_{\mathcal{R}_k}^{\mathcal{W}}} = \left(\begin{array}{cc|cc} \mathbf{I}_{3\times 3} & \mathbf{0}_{3\times 4} & \Delta t \cdot \mathbf{I}_{3\times 3} & \mathbf{0}_{3\times 3} \\ \mathbf{0}_{4\times 3} & \mathbf{Q}_{4\times 4} & \mathbf{0}_{4\times 3} & \mathbf{\Omega}_{4\times 3} \\ \hline \mathbf{0}_{6\times 7} & & \mathbf{I}_{6\times 6} & \end{array}\right) \tag{7}
$$

where $\mathbf{Q}$ and $\mathbf{\Omega}$ are defined as

$$
\begin{aligned} \mathbf{Q} &= \begin{pmatrix} q^R & -q^x & -q^y & -q^z \\ q^x & q^R & q^z & -q^y \\ q^y & -q^z & q^R & q^x \\ q^z & q^y & -q^x & q^R \end{pmatrix}, \text{ and} \\ \mathbf{\Omega} &= \begin{pmatrix} q_k^R & -q_k^x & -q_k^y & -q_k^z \\ q_k^x & q_k^R & -q_k^z & q_k^y \\ q_k^y & q_k^z & q_k^R & -q_k^x \\ q_k^z & -q_k^y & q_k^x & q_k^R \end{pmatrix} \cdot \frac{d\mathbf{q}}{d\omega} \cdot \Delta t. \end{aligned} \tag{8}
$$

with $\mathbf{q}_{\mathcal{R}_k}^{\mathcal{W}} = (q_k^R, q_k^x, q_k^y, q_k^z)^\top$ and $\exp(\omega^{\mathcal{R}}\Delta t) =$

$(q^R, q^x, q^y, q^z)^\top$. The measurement Jacobian is

$$\mathrm{H}^{\mathcal{R}_k} = \tag{9}$$

$$\left(\begin{array}{ccc|ccc} \frac{\partial \mathbf{h}_1^{\mathcal{R}_k}}{\partial \mathbf{r}_{\mathcal{R}}^{\mathcal{W}}} & \frac{\partial \mathbf{h}_1^{\mathcal{R}_k}}{\partial \mathbf{q}_{\mathcal{R}}^{\mathcal{W}}} & \mathbf{0}_{2\times 6} & \frac{\partial \mathbf{h}_1^{\mathcal{R}_k}}{\partial \mathbf{p}_k^{\mathcal{W}}} & \cdots & \mathbf{0}_{2\times 3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{h}_{N_f}^{\mathcal{R}_k}}{\partial \mathbf{r}_{\mathcal{R}}^{\mathcal{W}}} & \frac{\partial \mathbf{h}_{N_f}^{\mathcal{R}_k}}{\partial \mathbf{q}_{\mathcal{R}}^{\mathcal{W}}} & \mathbf{0}_{2\times 6} & \mathbf{0}_{2\times 3} & \cdots & \frac{\partial \mathbf{h}_{N_f}^{\mathcal{R}_k}}{\partial \mathbf{p}_k^{\mathcal{W}}} \\ \hline \frac{\partial \mathbf{h}_{(N_f+1)}^{\mathcal{R}_k}}{\partial \mathbf{r}_{\mathcal{R}}^{\mathcal{W}}} & \frac{\partial \mathbf{h}_{(N_f+1)}^{\mathcal{R}_k}}{\partial \mathbf{q}_{\mathcal{R}}^{\mathcal{W}}} & \mathbf{0}_{2\times 6} & \multicolumn{3}{c}{\multirow{3}{*}{$\mathbf{0}_{2N_a\times 3N_f}$}} \\ \vdots & \vdots & \vdots & & & \\ \frac{\partial \mathbf{h}_{(N_f+N_a)}^{\mathcal{R}_k}}{\partial \mathbf{r}_{\mathcal{R}}^{\mathcal{W}}} & \frac{\partial \mathbf{h}_{(N_f+N_a)}^{\mathcal{R}_k}}{\partial \mathbf{q}_{\mathcal{R}}^{\mathcal{W}}} & \mathbf{0}_{2\times 6} & & & \end{array}\right).$$

The first N_f rows are w.r.t. the features while the last N_a rows are w.r.t. the anchors. Using Equations (6) to (9), the dimensions of null spaces within one time segment when $N_f = 0, N_a \neq 0$ can be obtained: When $N_a \geq 3$, $\mathrm{Dim}(\mathcal{N}(\mathcal{Q}_{\mathrm{SOM}}(1))) = 0$ may hold, i.e. $\mathcal{Q}_{\mathrm{SOM}}(1)$ is full-rank. Thus, the system (3) is completely observable. Similarly, when $r \geq 2$ and $N_a \geq 1$, $\mathrm{Dim}(\mathcal{N}(\mathcal{Q}_{\mathrm{SOM}}(j))) = 0$ may hold, i.e. system is completely observable. □

According to Theorem 1, if a feature is tracked across 3 frames, the system composed of the camera motion and the feature may become observable, and the corresponding SOM full-rank. Degenerate conditions such as the point lying on the translation vector of a camera undergoing pure translation would fail to be observable (as would pure rotation). The degenerate conditions are typically of measure zero in the observation space. Tracking multiple features would guarantee observability for some subset of the tracked set. Under the observable condition for a feature, the value of a feature towards ego-motion estimation is reflected by the conditioning of the SOM. Thus, we define the **τ-temporal observability score** of a feature across τ **local frames,** $\tau \geq 2$ with the minimum singular value of SOM:

$$\psi(f, \tau) = \sigma_{\min}(\mathcal{Q}_{\mathrm{SOM}}(\tau|f)),$$

where at time k, $\mathcal{Q}_{\mathrm{SOM}}(\tau|f)$ is defined on the time segments $(k-\tau), (k-\tau+1), ..., k$.

This temporal observability score measures how constrained the SLAM estimate is w.r.t. the feature observation in the projective space, when considering the relative poses of the feature and camera over a recent period of time. The temporal nature of the measure is important because the SLAM estimate, in both the filtering and smoothing versions, is performed across time, with the current estimate affected by the previous estimate.

2.3. Rank-k Temporal Update of Observability Score

Computation of the τ-temporal observability score is efficient. Firstly, due to the sparse nature of the process matrix F, each subblock in $\mathcal{Q}$ can be computed iteratively with

$$\mathrm{HF}^n = \begin{pmatrix} \mathrm{H}_{1\sim 3} & \mathrm{H}_{4\sim 7}\mathbf{Q}^n & \mathrm{H}_{1\sim 3} n\Delta t & \mathrm{H}_{4\sim 7}\sum_{i=0}^{n-1}\mathbf{Q}^i\mathbf{\Omega} \end{pmatrix}$$

where $\mathrm{H}_{1\sim 3}$ denotes the matrix consist of column 1 to column 3 of matrix H. Secondly, the running temporal observability score of a feature can be computed efficiently with incremental SVD. Computation of the τ-temporal observability score is divided into the following phases:

1. In the first two frames that a feature is tracked, the observability cannot be full-rank. Build the SOM;
2. In frame three, the full rank condition of SOM may be satisfied. Compute SVD of the SOM;
3. From frame 4 to frame $\tau+1$ (in total τ time segments), for each new time segment a block of linear observability matrix is added to the SOM. Instead of computing SVD on the expanded SOM, perform a constant time rank-k update of the SVD [4], as per below.

   The SVD of $\mathcal{Q}_{\mathrm{SOM}}(j)$ is $USV^\top = \mathcal{Q}_{\mathrm{SOM}}(j)^\top$, where $S \in \mathbb{R}^{r\times r}$ with $r = 13$ (camera state). For the new row $\mathbf{a}^\top$, compute

   $$\mathbf{m} \triangleq \mathbf{U}^\top\mathbf{a};\ \mathbf{p} \triangleq \mathbf{a} - \mathbf{U}\mathbf{m};\ \mathbf{P} \triangleq \mathbf{p}/||\mathbf{p}||. \tag{10}$$

   Let

   $$\mathbf{K} = \begin{pmatrix} \mathbf{S} & \mathbf{m} \\ 0 & ||\mathbf{p}|| \end{pmatrix}. \tag{11}$$

   Diagonalize $\mathbf{K}$ as $\mathbf{U}'^\top\mathbf{K}\mathbf{V}' = \mathbf{S}'$ and update

   $$[\mathcal{Q}_{\mathrm{SOM}}(j)^\top|a] = ([\mathbf{U}\ \mathbf{P}]\mathbf{U}')\mathbf{S}'([\bar{\mathbf{V}}\ \mathbf{Q}]\mathbf{V}')^\top \tag{12}$$

   where $\bar{\mathbf{V}}^\top = [\mathbf{V}^\top, \mathbf{0}]$, $\mathbf{Q} = [0, \cdots, 0, 1]^\top$. Diagonalization of $\mathbf{K}$ takes $\mathcal{O}(r^2)$ [16].

   Expanding the SOM with more time segments results in adding $2r$ new rows into SOM. Each new row requires a rank-1 update, leading to rank-$2r$ update for the whole SOM.
4. After frame $\tau+1$, for each new frame, update the SOM by replacing the subblock from the oldest time segment with the linear observability matrix of the current time segment. For example, let SOM at time k be $\mathcal{Q}_{\mathrm{SOM}}^{(k)}(\tau) = \left[\mathcal{Q}_{k-\tau+1}^\top|\mathcal{Q}_{k-\tau+2}^\top|\cdots|\mathcal{Q}_k^\top\right]^\top$, then at time $k+1$, $\mathcal{Q}_{\mathrm{SOM}}^{(k+1)}(\tau) = \left[\mathcal{Q}_{\mathbf{k+1}}^\top|\mathcal{Q}_{k-\tau+2}^\top|\cdots|\mathcal{Q}_k^\top\right]^\top$.

   Computing the SVD of $\mathcal{Q}_{\mathrm{SOM}}^{(k+1)}(\tau)$ given the SVD of $\mathcal{Q}_{\mathrm{SOM}}^{(k)}(\tau)$ can also be done with a rank-$2r$ update similar to phase 3. Let row b be replaced by row vector c in this case, by setting $a = (c-b)^\top$, the updated SVD is generated via (10)-(12).

After updating the τ-temporal observability scores of the features and ranking them, the top K_a features over a selected threshold are upgraded to be anchors. If the anchor set has less than $(K_a - 2)$ elements passing the threshold test, then additional features will need to be added to complete the anchor set.

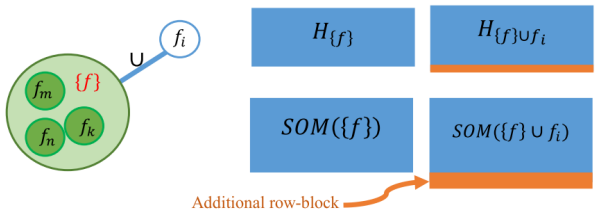

Figure 3. In spatial grouping, selecting one more feature as anchor results in an additional row-block in the measurement Jacobian, which further expands the SOM.

3. Submodular Learning for Feature Grouping

When needed, the group completion step selects more features as anchors by maximizing the minimum singular value of SOM over the selected features. Upgrading a feature to be an anchor will expand the dimension of F and H in Equation (6)-(9), resulting in additional rows in SOM.

The group completion problem can be formulated as follows: Let X be the SOM of the features with high observability score, $X \in \mathbb{R}^{n\times m}, n \geq m$. Adding a feature results in adding a row-block R_k to the SOM as in Fig. 3. Denote the set of all candidate row-blocks as $\mathbf{R} = \{R_1, R_2, ..., R_K\}$, $R_k \in \mathbb{R}^{n'\times m}$. Finding K^* features which form the most observable SLAM subsystem is equivalent to finding a subset of the candidate rows that maximize the minimum singular value of the augmented matrix

$$\mathbf{R}^* = \underset{\mathbf{R}^*\subseteq\mathbf{R},|\mathbf{R}^*|=K^*}{\operatorname{argmax}} \sigma_{min}\left(\left[X^\top|R_1^{*\top}|R_2^{*\top}|...|R_{K^*}^{*\top}\right]^\top\right)$$

Such a combinatorial optimization problem is NP-hard. However, the problem has nice submodular properties.

Definition 1. *[23] **(Approximate submodularity)** A set function $F: 2^{\mathbf{V}} \mapsto \mathbb{R}$ is approximately submodular if for $\mathbf{D} \subset \mathbf{D}' \subset \mathbf{V}$ and $v \in \mathbf{V} \setminus \mathbf{D}'$*

$$F(\mathbf{D}\cup\{v\}) - F(\mathbf{D}) \geq F(\mathbf{D}'\cup\{v\}) - F(\mathbf{D}') - \varepsilon \quad (13)$$

Theorem 2. *When $\mathbf{X}\cap\mathbf{R} = \emptyset$, the set function $F_{\sigma_{min}}(\cdot): 2^{\mathbf{X}\cup\mathbf{R}} \mapsto \mathbb{R}$ is approximately submodular,*

$$F_{\sigma_{min}}(\mathbf{X}\cup\mathbf{R}^*) = \sigma_{min}\left(\left[X^\top|R_1^{*\top}|R_2^{*\top}|...|R_{K^*}^{*\top}\right]^\top\right). \quad (14)$$

The proof requires the following two lemmas.

Lamma 3. *[3] **(Concavity of min eigenvalue function)** For any real symmetric matrix $G \in \mathbb{R}^{m\times m}$, let $f(G) \triangleq \lambda_{min}(G)$, $f(G)$ is a concave function of G.*

Lamma 4. *[13] **(Eigenvalues of sum of two matrices)** Let A, B, C be Hermitian n by n matrices, denote the eigenvalues of A by $\alpha: \alpha_1 \geq \alpha_2 \geq ... \geq \alpha_n$, and similarly write β and γ for eigenvalues of B and C, then:*

$$\gamma_{i+j-1} \leq \alpha_i + \beta_j \quad \text{whenever } i+j-1 \leq n. \quad (15)$$

Proof. **(Theorem 2)** WLOG consider the two row-blocks R_1 and R_2 from $\mathbf{R}$. Denote the Gram matrices $G_\circ$ as:

$$G_X = X^\top X,\ G_{R_1} = R_1^\top R_1, \text{ and } G_{R_2} = R_2^\top R_2.$$

Also define the augmented Gram matrices as

$$G_{XR} = \left(X^\top|R^\top\right)\cdot\begin{pmatrix}X\\R\end{pmatrix}$$

It holds that $G_{XR_1} = G_X + G_{R_1}$, $G_{XR_2} = G_X + G_{R_2}$, $G_{XR_1R_2} = G_{XR_1} + G_{R_2}$, and $G_{XR_2R_1} = G_{XR_2} + G_{R_1}$. Let the minimum eigenvalue of G_X be $\lambda_{min}(G_X) \equiv \lambda_m(G_X)$, the maximum eigenvalue be $\lambda_{max}(G_X) \equiv \lambda_1(G_X)$. Since X is a real matrix, $\lambda_{min}(G_X) = \sigma_{min}^2(X)$, and likewise for the augmented matrices. From Lemma 3,

$$\begin{aligned}\lambda_{min}(G_{XR1}) &= \lambda_{min}(G_X + G_{R1})\\ &\geq (\lambda_{min}(G_X) + \lambda_{min}(G_{R1}))\\ &\geq \lambda_{min}(G_X)\end{aligned} \quad (16)$$

Thus, $F_{\sigma_{min}}(\mathbf{X}\cup\{R_1\}) \geq F_{\sigma_{min}}(\mathbf{X})$. From Lemma 4, and the fact that the Gram matrices are real-symmetric and hence Hermitian, the following holds:

$$\begin{aligned}\lambda_{min}(G_{XR1R2}) &= \lambda_{m+1-1}(G_{XR1R2})\\ &\leq \lambda_m(G_{XR2}) + \lambda_1(G_{R1})\end{aligned} \quad (17)$$

Combining (16) and (17),

$$\begin{aligned}&\lambda_{min}(G_X) + \lambda_{min}(G_{XR1R2})\\ &\quad\leq \lambda_{min}(G_{XR1}) + \lambda_{min}(G_{XR2}) + d_\rho(R_1),\end{aligned}$$

where $d_\rho(R_1) = \lambda_{max}(R_1) - \lambda_{min}(R_1)$. Similarly,

$$\begin{aligned}&\lambda_{min}(G_X) + \lambda_{min}(G_{XR1R2})\\ &\quad\leq \lambda_{min}(G_{XR1}) + \lambda_{min}(G_{XR2}) + d_\rho(R_2)\end{aligned}$$

The tighter bound is:

$$\begin{aligned}&\lambda_{min}(G_X) + \lambda_{min}(G_{XR1R2})\\ &\leq \lambda_{min}(G_{XR1}) + \lambda_{min}(G_{XR2}) + \min(d_\rho(R_1), d_\rho(R_2)).\end{aligned}$$

This leads to

$$\begin{aligned}&F_{\sigma_{min}}(\mathbf{X}\cup\{R_1\}) + F_{\sigma_{min}}(\mathbf{X}\cup\{R_2\})\\ &\quad\geq F_{\sigma_{min}}(\mathbf{X}) + F_{\sigma_{min}}(\mathbf{X}\cup\{R_1\}\cup\{R_2\})\\ &\qquad - \min(d_\rho(R_1), d_\rho(R_2)).\end{aligned} \quad (18)$$

When $\mathbf{X}\cap\mathbf{R} = \emptyset$, $F_{\sigma_{min}}(\cdot)$ is approximately submodular, with the bound $\varepsilon = \max(d_\rho(R_k))$, $\forall R_k \in \mathbf{R}$. □

Theorem 2 means that a greedy algorithm will be near-optimal. The simplest greedy algorithm outline in Algorithm 1 identifies the group completion in the cardinality deficient case with a complexity of $\mathcal{O}(K^*Kn')$ (when using incremental SVD). The near-optimality bound is

Theorem 1. *[23]. Let $\mathbf{A_G}$ be the set of the first K^* elements chosen by Algorithm 1, and let $OPT = \max_{\mathbf{A}\subset\mathbf{R},|\mathbf{A}|=K^*} F_{\sigma_{min}}(\mathbf{X}\cup\mathbf{A})$. Then*

$$F_{\sigma_{min}}(\mathbf{A_G}) \geq \left(1 - \left(\frac{K^*-1}{K^*}\right)^{K^*}\right)(OPT - K^*\varepsilon) \quad (19)$$

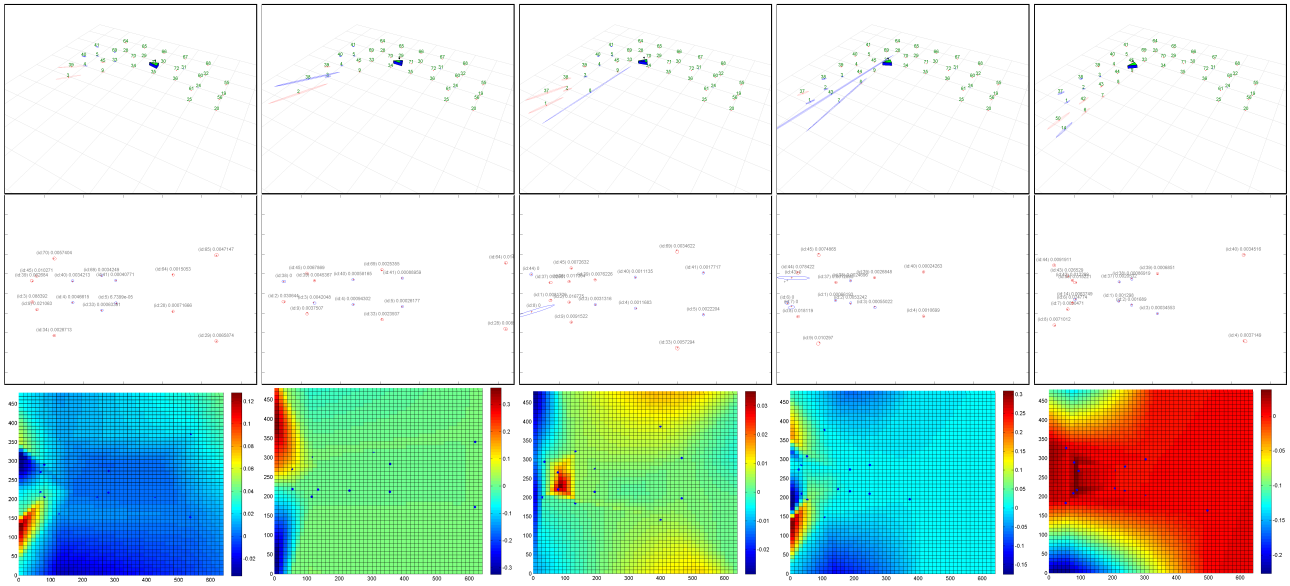

Figure 4. Simulated scenario #1 for ego-motion estimation experiment. Results shown have 1.0 pixel measurement standard deviation and $K_a = 10$. Row 1: reconstructed maps at time steps when camera is performing circular movement; features are depicted with estimated mean and covariance; points in red are selected as anchors. Row 2: corresponding camera frames with observability scores shown for all measurements. Row 3: interpolated maps of observability score on image plane showing how it changes during the motion.

Algorithm 1: Submodular learning for feature grouping.

Data: $X \in \mathbb{R}^{n \times m}, n \geq m,$
$\mathbf{R} = \{R_1, R_2, ..., R_K\},\ R_k \in \mathbb{R}^{1 \times m}, K^*$
Result: $\mathbf{R}^*,\ |\mathbf{R}^*| = K^*$

$\mathbf{R}^* \leftarrow \emptyset$;
while $|\mathbf{R}^*| < K^*$ **do**
  $R^* \leftarrow \arg\max_{R^* \in \mathbf{R}} F_{\sigma_{min}}(\mathbf{X} \cup \{R^*\})$;
  $\mathbf{R}^* \leftarrow \mathbf{R}^* \cup \{R^*\}$;
  $\mathbf{R} \leftarrow \mathbf{R} \setminus \{R^*\}$;

4. Evaluation

4.1. Integration into SLAM

The proposed method is complementary to various sequential SLAM algorithms [11, 7, 22, 21]. It provides a ranking of features which can be used in different phases:

- **Ego-motion estimation.** After data-association but prior to post-optimization, the Good Features method can be used to select a subset of features from the best matched measurements, so that both the data-association scores and the observability scores are considered. Localization is performed using only the subset, while the mapping is performed on the whole feature set based on the localization results (acting as external input).
- **Data-association.** The observability scores can be used in some data-association processes. For example, in 1-Point RANSAC method, for each iteration the features with high observability scores are used to partially update the model, which is then used to retrieve the inlier set.

4.2. Experiments for Ego-motion estimation

Evaluation of the proposed method for the ego-motion estimation phase focuses on the estimation accuracy. Therefore this experiment will isolate the data association error from the localization error such that the SLAM accuracy is only affected by the selection of anchors. Precisely benchmarking the SLAM accuracy is a difficult task, because most of the publicly available datasets do not provide exact ground truth and perfect data association. The usual SLAM baseline for evaluating accuracy is a global optimization, usually bundle adjustment [2, 35]. However, these data-driven baseline methods are not actual ground truth.

Experimental Scenarios. To perform controlled experiments for accuracy evaluation, we use camera motion and observation simulation modules from software in [31] which assumes perfect data association, but implements the SLAM estimation process. Two scenarios are simulated. The simulated environment is of dimension $12m \times 12m$ with 72 landmarks forming a square. Two scenarios are tested. In the first one, the robot performs circular trajectory as in Row 1, Fig. 4. The second scenario simulates a more cluttered scene. The robot moves away from the landmarks while performing slight rotation as shown in Fig. 5.

Experiment Setup and Comparison. In each time step, ego-motion estimation is performed with an Extended

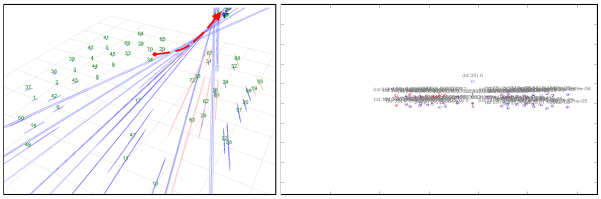

Figure 5. Simulated scenario #2 for ego-motion estimation experiment. Left: reconstructed map with robot trajectory shown in the thick red curve. Right: corresponding camera frame.

Kalman Filter **only with the anchors**, while the features are estimated based on the ego-motion estimate. Experiments are performed with different levels of observation noise and anchor set sizes. We tested the configurations with standard deviation of observation noise of $0.5,\ 1.0,\ 1.5,\ 2.0,\ 2.5$ pixels under Gaussian noise, and maximum anchors sets of $K_a = 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 10,\ 12$. The temporal parameter $\tau = 5$ is used in our method. The threshold for observability score is 0.003. In cases with less than $K_a - 2$ strongly observable features, at most 2 more features are added via spatial grouping. The baseline state-of-the-art method uses information gain for feature selection [20]. The same ego-motion estimation and mapping scheme is applied on both methods. Due to the randomized effects from the noise simulation, 15 experiments are run per configuration.

Metrics. Localization accuracy is evaluated by the cumulative translation errors and cumulative orientation errors. Let $\Delta\mathbf{r}^{\mathcal{W}}_{\mathcal{R}_k}$ be the translation error at time k, $\Delta\theta^{\mathcal{W}}_{\mathcal{R}_k}$ be the orientation error in Euler angles, $\sum_k ||\Delta\mathbf{r}^{\mathcal{W}}_{\mathcal{R}_k}||_2$ and $\sum_k ||\Delta\mathbf{r}^{\mathcal{W}}_{\mathcal{R}_k}||_\infty$ are used for evaluating cumulative translation errors, and accordingly $\sum_k ||\Delta\theta^{\mathcal{W}}_{\mathcal{R}_k}||_2$ and $\sum_k ||\Delta\theta^{\mathcal{W}}_{\mathcal{R}_k}||_\infty$ for cumulative orientation errors. The average value of the 15 runs are used as the final evaluation result for each configuration.

Results. The evaluation results are shown in Fig. 6 and Fig. 7. For the interest of space, configurations of #Anchors $\in \{3, 4, 5, 10\}$ are displayed for scenario #1 to highlight both the extreme cases and saturated cases, and #Anchors $\in \{3, 4, 8, 10\}$ for the more cluttered scenario #2. Our method outperforms the information gain based method in 92.5% (37/40) cases for translation and 82.5% (33/40) for orientation in scenario #1; 85% (34/40) for translation and 95% (38/40) for orientation in scenario #2. These ratios are the same for both l_2-norm and l_∞-norm metrics.

4.3. Experiments for Data-association

The proposed method is tested in data-association with real scenes and via modification of the baseline SLAM system (1-Point RANSAC) from [7]. For data-association, the features are first matched with individual compatibility. Then in each iteration of 1-Point RANSAC, one feature measurement is selected randomly to partially update the localization, which further generates a hypothesis to retrieve the inlier set. The maximum supported hypothesis is used as the data-association results. The Good Features modification changes selection of the feature for hypothesis generation such that strongly observable features are selected.

Figure 6. Results of simulation scenario #1 with cumulative translation errors and cumulative orientation errors. "ObsStd "stands for the standard deviation of observation noise in pixel units.

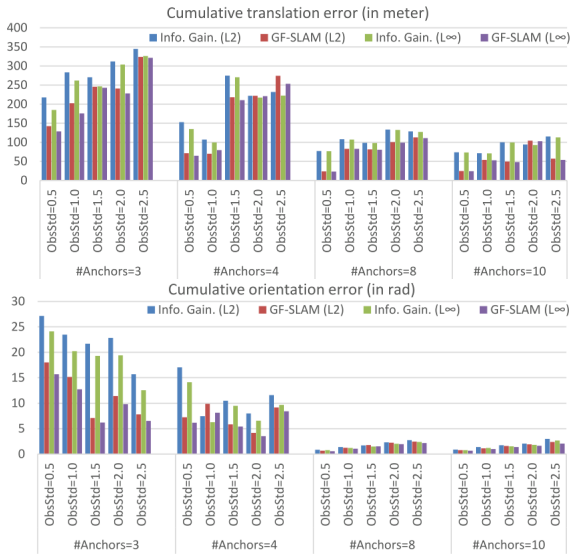

Figure 7. Results of simulation scenario #2.

Dataset. For the purpose of evaluating the effect of temporal parameter, we collected videos under smooth motion and highly dynamic motion respectively. We use 3 videos for each type of motion respectively. The videos are collected in 640×480 resolution and 40 fps frame rate. Each video clip has about 2300 frames.

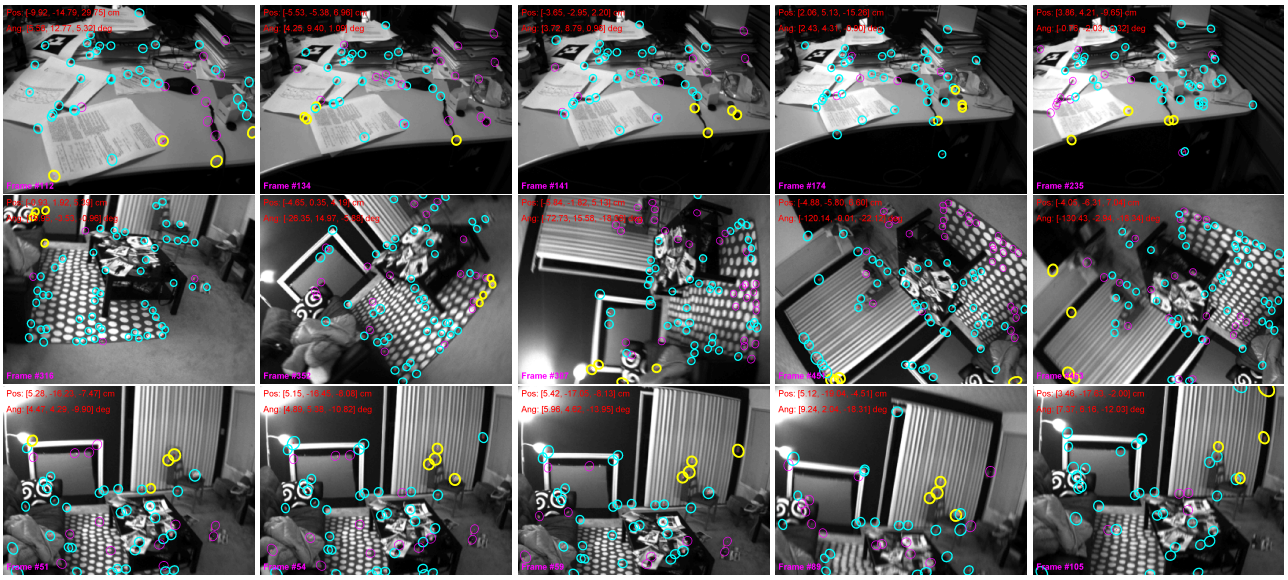

Figure 8. Example frames from data-association experiment. The strongly observable features are illustrated in yellow, retrieved inlier set is in cyan, and the outlier set is in purple. Row 1: camera is moving away from the desktop. Row 2: camera is rotating w.r.t. the optical axis. Row 3: camera is rotating w.r.t. the x axis of camera.

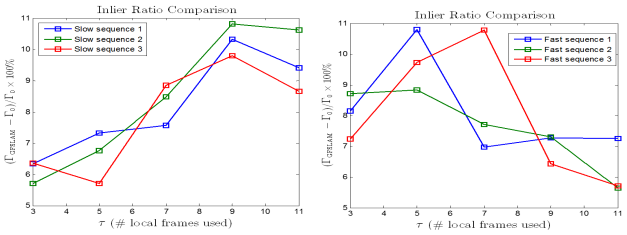

Figure 9. Relative improvements of inlier ratios versus [7].

Experiment Setup and Comparison. The code was written in C++ with OpenCV and Armadillo following the pipeline described in [7]. The experiments are run on a 2.7GHz 8-core PC with 16GB RAM. For our method, the strongly observable features quantity parameter is set to K_a, which are then used to generate the data-association hypothesis. We tested our method with temporal parameter $\tau \in \{3, 5, 7, 9, 11\}$. Some example frames under three motion segments are shown in Fig. 8.

Metrics. We evaluate data-association results by comparing average inlier ratios of the maximum supported data-association hypothesis. The inlier ratio is defined as $\Gamma = \#inlier/(\#inliers + \#outliers)$.

Results. The relative improvements of the inlier ratios from our good features for SLAM method (denoted as Γ_{GFSLAM}) over that from [7] (denoted as Γ_0) are shown in Fig. 9. Our method outperforms [7] in all the datasets by at least ≈5.5%. For the slow motion, the inlier ratio of our method has the peak value with $\tau \in [9, 11]$. For the fast motion, the peak value is at about $\tau \in [5, 7]$. Some statistics of execution time are reported in Table 1. Our method has a slightly lower average frame rate due to the computation overhead for computing the observability scores. However, our method is more stable in terms of both standard deviation and max/min values. Moreover, the execution time of our method can be further improved by parallelizing the computation of observability scores for different features.

Frame rates (fps)	Mean	Std.	Min	Max
[7]	52.58	9.25	12.26	72.55
GF-SLAM	48.14	5.59	25.81	65.54

Table 1. Statistics of execution time.

5. Conclusion

We presented a new method for selecting the features in visual SLAM process which provides the best values for SLAM estimation. The feature selection criterion based on temporal observability is proposed via analysis of the visual SLAM problem from a control systems view. We further develop efficient computation methods for temporally updating the score via incremental SVD. A greedy algorithm for group completion, in the case of insufficient high-observability features, is also presented and justified. The Good Features method performs competitively with respect to the state-of-the-art methods in terms of localization accuracy and data-association inlier ratios.

Acknowledgment

This work was supported by the AFRL research award FA9453-13-C-0201.

References

[1] J. Andrade-Cetto and A. Sanfeliu. The effects of partial observability when building fully correlated maps. *IEEE Transactions on Robotics*, 21(4):771–777, 2005.

[2] J. Balzer and S. Soatto. CLAM: Coupled localization and mapping with efficient outlier handling. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1554–1561, 2013.

[3] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2009.

[4] M. Brand. Fast low-rank modifications of the thin singular value decomposition. *Linear Algebra and its Applications*, 415(1):20–30, 2006.

[5] L. Carlone, P. F. Alcantarilla, H.-P. Chiu, Z. Kira, and F. Dellaert. Mining structure fragments for smart bundle adjustment. In *British Machine Vision Conference*, 2014.

[6] M. Chli and A. J. Davison. Active matching. In *European Conference on Computer Vision*, pages 72–85, 2008.

[7] J. Civera, O. G. Grasa, A. J. Davison, and J. Montiel. 1-Point RANSAC for extended Kalman filtering: Application to real-time structure from motion and visual odometry. *Journal of Field Robotics*, 27(5):609–631, 2010.

[8] A. Dame, V. A. Prisacariu, C. Y. Ren, and I. Reid. Dense reconstruction using 3D object shape priors. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[9] A. Davison. Active search for real-time vision. In *IEEE International Conference on Computer Vision*, volume 1, pages 66–73, 2005.

[10] A. J. Davison, Y. G. Cid, and N. Kita. Real-time 3D SLAM with wide-angle vision. In *IFAC/EURON Symposium on Intelligent Autonomous Vehicles*, 2004.

[11] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse. MonoSLAM: Real-time single camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1052–1067, 2007.

[12] J. Engel, T. Schöps, and D. Cremers. LSD-SLAM: Large-Scale Direct monocular SLAM. In *European Conference on Computer Vision*, pages 834–849. 2014.

[13] W. Fulton. Eigenvalues, invariant factors, highest weights, and schubert calculus. *Bulletin of the American Mathematical Society*, 37(3):209–249, 2000.

[14] P. W. Gibbens, G. M. Dissanayake, and H. F. Durrant-Whyte. A closed-form solution to the single degree of freedom simultaneous localisation and map building (SLAM) problem. In *IEEE Conference on Decision and Control*, volume 1, pages 191–196, 2000.

[15] D. Goshen-Meskin and I. Bar-Itzhack. Observability analysis of piece-wise constant systems. I. theory. *IEEE Transactions on Aerospace and Electronic Systems*, 28(4):1056–1067, 1992.

[16] M. Gu and E. Stanley C. A stable and fast algorithm for updating the singular value decomposition. *Technical Report YALEU/DCS/RR-966, Department of Computer Science*, 1993.

[17] A. Handa, M. Chli, H. Strasdat, and A. Davison. Scalable active matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1546–1553, 2010.

[18] G. P. Huang, A. I. Mourikis, and S. I. Roumeliotis. Observability-based rules for designing consistent EKF SLAM estimators. *International Journal of Robotics Research*, 29(5):502–528, 2010.

[19] H. Joo, H. Soo Park, and Y. Sheikh. MAP visibility estimation for large-scale dynamic 3D reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1122–1129, 2014.

[20] M. Kaess and F. Dellaert. Covariance recovery from a square root information matrix for data association. *Robotics and Autonomous Systems*, 57(12):1198–1210, 2009.

[21] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert. iSAM2: Incremental smoothing and mapping using the bayes tree. *International Journal of Robotics Research*, 31:217–236, 2012.

[22] G. Klein and D. Murray. Parallel tracking and mapping for small AR workspaces. In *IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 225–234, 2007.

[23] A. Krause, A. Singh, and C. Guestrin. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9:235–284, 2008.

[24] K. W. Lee, W. S. Wijesoma, and I. G. Javier. On the observability and observability analysis of SLAM. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3569–3574, 2006.

[25] R. M. Murray, Z. Li, S. S. Sastry, and S. S. Sastry. *A mathematical introduction to robotic manipulation*. CRC Press, 1994.

[26] A. Nemra and N. Aouf. Robust airborne 3D visual simultaneous localization and mapping with observability and consistency analysis. *Journal of Intelligent and Robotic Systems*, 55(4):345–376, 2009.

[27] R. A. Newcombe and A. J. Davison. Live dense reconstruction with a single moving camera. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1498–1505, 2010.

[28] L. D. L. Perera and E. Nettleton. On the nonlinear observability and the information form of the slam problem. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2061–2068, 2009.

[29] M. Pizzoli, C. Forster, and D. Scaramuzza. REMODE: Probabilistic, monocular dense reconstruction in real time. In *IEEE International Conference on Robotics and Automation*, pages 2609–2616, 2014.

[30] J. Shi and C. Tomasi. Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1994.

[31] J. Sola, T. Vidal-Calleja, J. Civera, and J. M. M. Montiel. Impact of landmark parametrization on monocular EKF-SLAM with points and lines. *International Journal of Computer Vision*, 97(3):339–368, 2012.

[32] B. Southall, B. F. Buxton, and J. A. Marchant. Controllability and observability: Tools for Kalman filter design. In *British Machine Vision Conference*, pages 1–10, 1998.

[33] A. Vedaldi, H. Jin, P. Favaro, and S. Soatto. KALMANSAC: Robust filtering by consensus. In *IEEE International Conference on Computer Vision*, pages 633–640, 2005.

[34] T. Vidal-Calleja, M. Bryson, S. Sukkarieh, A. Sanfeliu, and J. Andrade-Cetto. On the observability of bearing-only SLAM. In *IEEE International Conference on Robotics and Automation*, pages 4114–4119, 2007.

[35] G. Zhang, X. Qin, W. Hua, T.-T. Wong, P.-A. Heng, and H. Bao. Robust metric reconstruction from challenging video sequences. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.