Good Features to Track for Visual SLAM

Guangcong Zhang[1], Patricio A. Vela[1]
[1]School of ECE, and Institute of Robotics and Intelligent Machines, Georgia Institute of Technology.

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. Conventionally, a fully data-driven and randomized process like RANSAC is used to select the valuable features by retrieving the inlier set [5]. Information gain [3, 4] has also been a popular criterion for such a selection, will maximize the uncertainty reduction for both the camera pose and landmark positions. Recent research efforts have sought more systematic criteria. [2] exploits the co-visibility of features by cameras to select the best subset of points, but it requires the complete structure of features-camera graph as a priori knowledge.

This paper presents a novel method for selecting a subset of features that are of high utility for localization in the SLAM/SfM estimation process, by examining the **observability of SLAM**. Being complimentary to the estimation process, it easily integrates into existing SLAM systems.

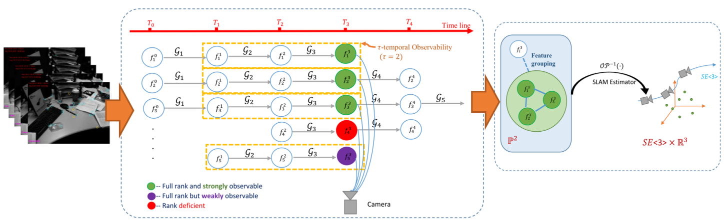

Figure 1: Overview of our approach.

Overview. As depicted in Figure 1, in a time step, we first examine the rank conditions for features, i.e. whether the feature is **completely observable** to the SLAM system. If rank condition is satisfied (depicted in green/purple), the **τ-temporal observability score** is evaluated, and features with high scores are selected (depicted in green). When needed, feature grouping with a submodular learning scheme is applied to collect more good features.

System Modeling. We measure the system observability by first modeling the a $SE\langle 3\rangle$ SLAM as a **Piece-wise Linear System (PWLS)**. Assume the SLAM system has N_f features and N_a anchors. For the k-th time segment $\mathcal{T}_k \equiv [t_k, t_{k+1})$ (from time k to time $k+1$), the dynamics are

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

Under smooth motion assumption, system at $\mathcal{T}_k$ is linearized as a **PWLS** [6].

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

System Observable Conditions. A PWLS is **completely observable** iff Total Observability Matrix (TOM) is **full-rank**, but computing TOM is expensive. Lemma 1 provides a proxy to examine full rank conditions.

Lamma 1. *[6] For PWLS, when $\mathcal{N}(\mathcal{Q}_j) \subset \mathcal{N}(F_j)$, the stripped Observability Matrix (SOM). $\mathcal{Q}_{\text{SOM}}(j) = \left[\mathcal{Q}_1^\top | \mathcal{Q}_2^\top | \cdots | \mathcal{Q}_j^\top\right]^\top$. has the same nullspace as TOM, i.e. $\mathcal{N}(\mathcal{Q}_{\text{SOM}}(j)) = \mathcal{N}(\mathcal{Q}_{\text{TOM}}(j))$. $\mathcal{Q}_j$ is the linear observability matrix for time segment j.*

Our work proves the completely observable conditions of $SE\langle 3\rangle$ SLAM.

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

System Observability Measure. We define the **τ-temporal observability score** of a feature across τ local frames, $\tau \geq 2$ as:

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

where at time k, $\mathcal{Q}_{\text{SOM}}(\tau|f)$ 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.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

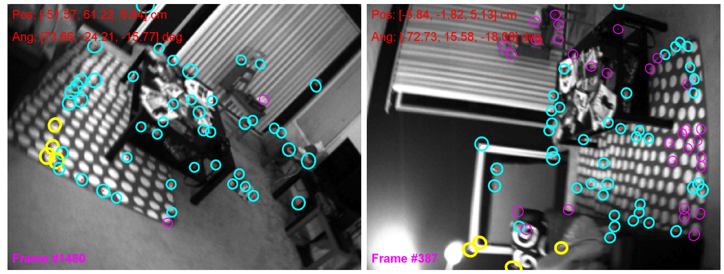

Figure 2: Measurements selected (highlighted in yellow) by considering observability scores. In the example the camera is mostly rotating w.r.t. the optical axis.

Rank-k Temporal Update of Observability Score. Computation of the τ-temporal observability score is efficient. Firstly, each subblock in $\mathcal{Q}$ is computed iteratively with $\mathrm{HF}^n = \left(\mathrm{H}_{1\sim 3} \quad \mathrm{H}_{4\sim 7}\mathbf{Q}^n \quad \mathrm{H}_{1\sim 3}n\Delta t \quad \mathrm{H}_{4\sim 7}\sum_{i=0}^{n-1}\mathbf{Q}^i\Omega\right)$.

Secondly, the running temporal observability score is computed efficiently with incremental SVD [1] with 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, with a **constant time rank-k update** of the SVD;
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, again by a **rank-k update**.

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. Adding a feature results in adding a row-block R_k to the SOM. 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

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

This is an NP-hard problem. Our work proves that the objective function is **approximately submodular**.

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

Thus, a **greedy** selection algorithm gives a *near-optimal* solution [7].

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

[2] Luca Carlone and et. al. Mining structure fragments for smart bundle adjustment. In *BMVC*, 2014.

[3] Margarita Chli and Andrew J Davison. Active matching. In *ECCV*, pages 72–85, 2008.

[4] A.J. Davison. Active search for real-time vision. In *ICCV*, volume 1, pages 66–73, 2005.

[5] Andrew J Davison and et. al. MonoSLAM: Real-time single camera SLAM. *T-PAMI*, 29(6):1052–1067, 2007.

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

[7] Andreas Krause and et. al. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9:235–284, 2008.