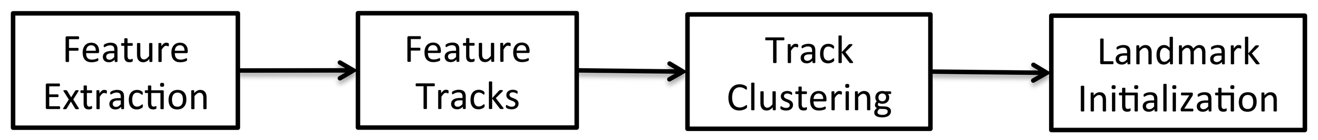
Initialization in Monocular SLAM

Martin Skoglund*, Zoran Sjanic* †, Thomas B. Schön* and Fredrik Gustafsson*
*Linköping University †Saab Aeronautics

Main idea

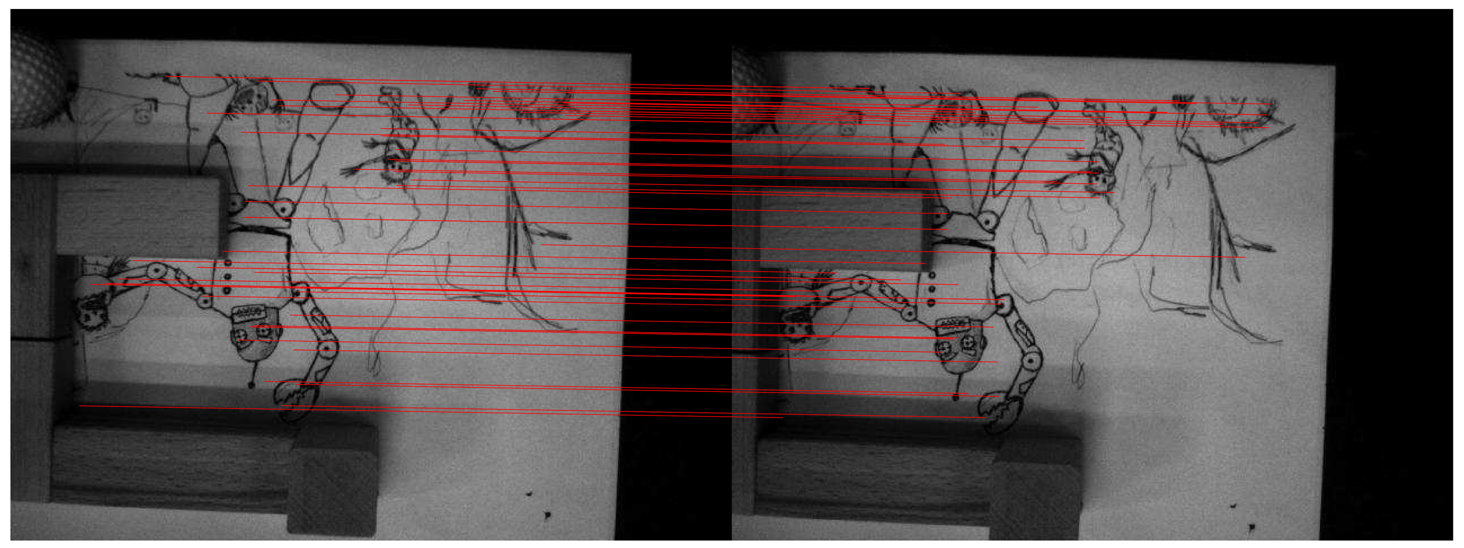
We propose an initialization method for monocular SLAM using sequential data. Feature tracks are extracted from SIFT correspondence pairs in the image sequence. These tracks are clustered based upon their average feature descriptor in order to find loop closure candidates. The clustered featured tracks are then used to initialize 3D point landmarks. The initialization can easily be used for inference with other sensors e.g., inertial.



Camera-only initialization procedure

Introduction

The correspondence problem is fundamental in applications such as Structure from Motion (SfM) and Bundle Adjustment (BA). With sequential data, local features, are tracked in order to find the relative displacements of the cameras and the positions of the tracked features. While locally consistent correspondences are rather easy to obtain, consistent loop closures are more difficult.



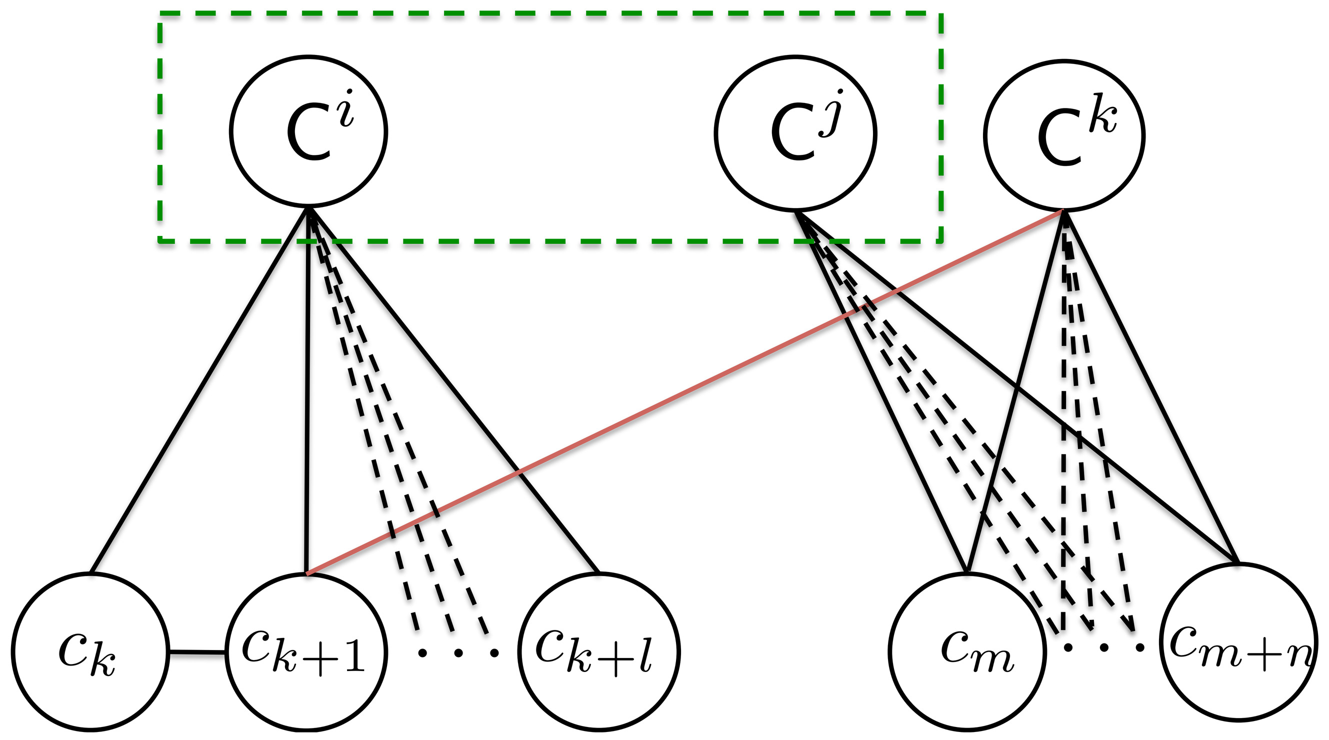
Feature Matching

We use feature vectors $\boldsymbol{f}_t$ from Scale-Invariant Feature Transform (SIFT) to find pairwise matches in the image sequence from the descriptors only. This is done by constructing a matrix containing the pairwise distances of the features in each consecutive image pair $\mathsf{G}^{ij} = -||\boldsymbol{f}_t^i - \boldsymbol{f}_{t+1}^j||^{-2}$. The matching problem then consist of assigning a subset of measurements from image t , y_t^i, $i \in N$, to a subset of measurements from image $t+1$, y_{t+1}^j, $j \in M$ such that each measurement gets assigned to exactly one, unique, other measurement. These assignments are encoded by binary correspondence variables $c^{ij} \in \{0,1\}$ and each assignment is associated with the matching cost G^{ij}.

$$\begin{aligned} \arg\min_{c^{ij}} \; & \sum_{i=1}^{N}\sum_{j=1}^{M} \mathsf{G}^{ij}\, c^{ij} \\ \text{subject to} \; & \sum_{i=1}^{N} c^{ij} \leq 1,\; j \in M \\ & \sum_{j=1}^{M} c^{ij} \leq 1,\; i \in N \\ & c^{ij} \in \{0,1\} \end{aligned}$$

It is important to note that the constraint matrix is unimodular because then the linear relaxation, $0 \leq c^{ij} \leq 1$, attains the same integral optimum as the original problem.

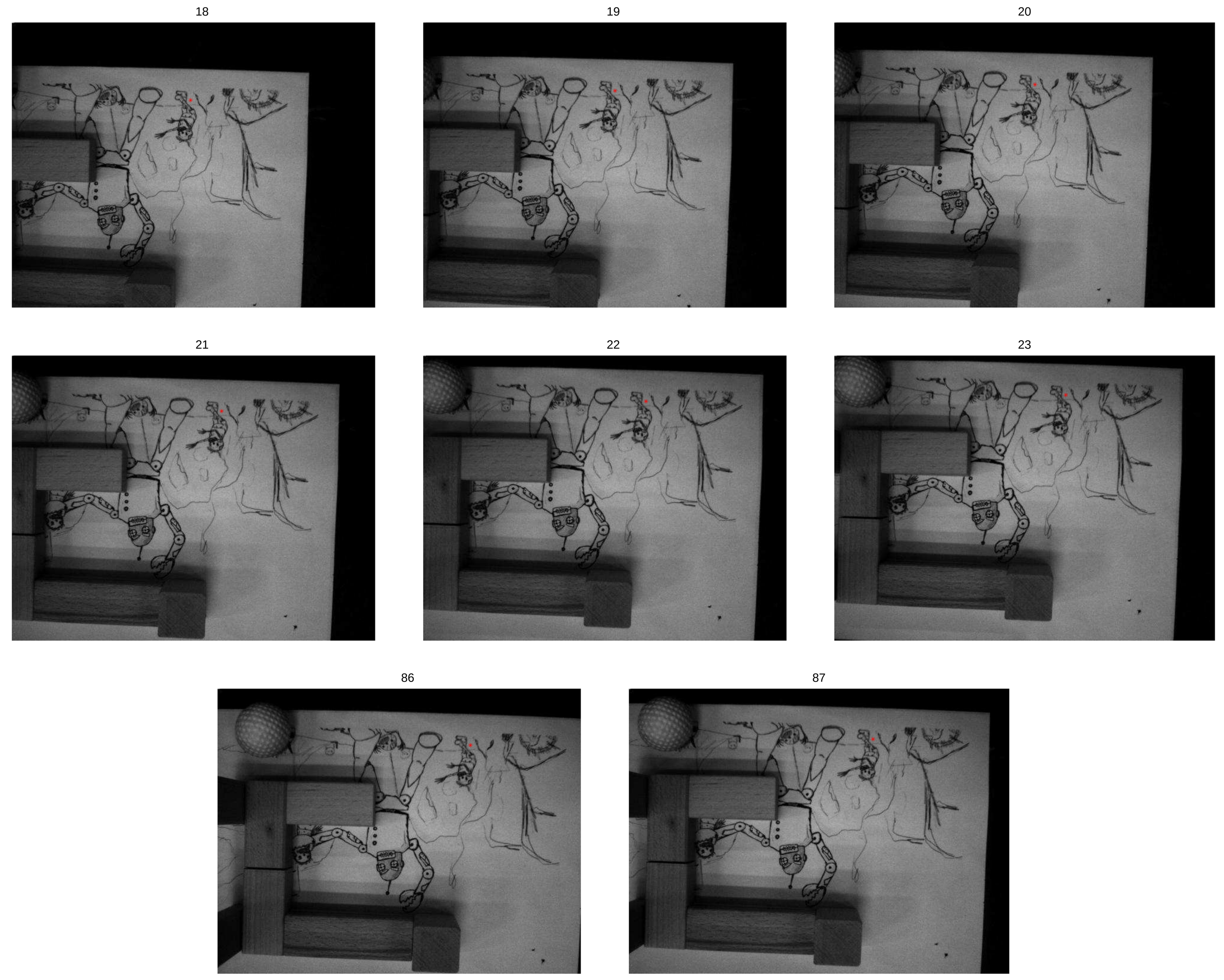
Clustering Feature Tracks



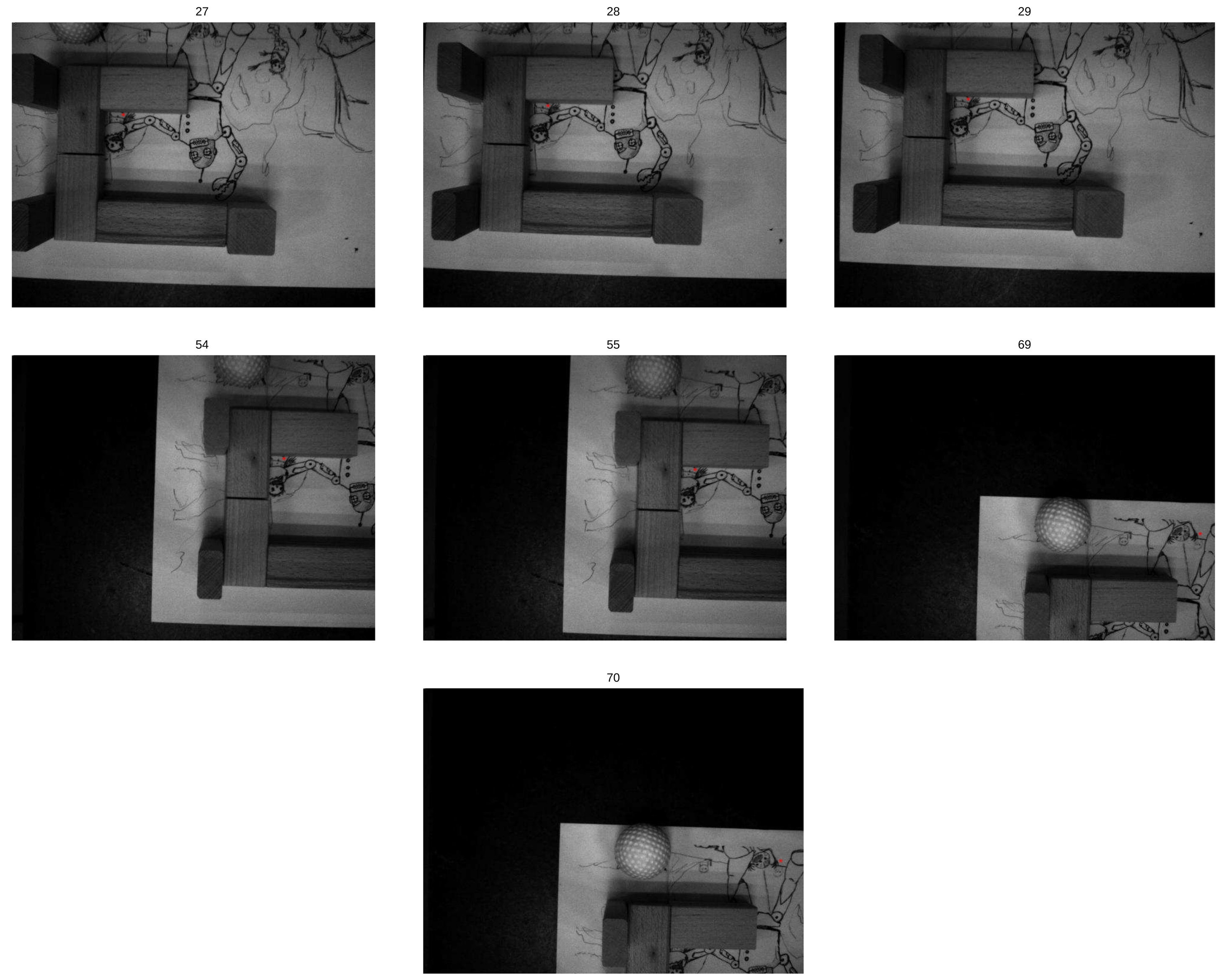
Feature tracks are defined as a time sequence of pairwise matching feature correspondences $\mathsf{C} = [c_k, c_{k+1}, \ldots, c_{k+l}]$. A valid clusters of feature tracks have only time-disjoint tracks, i.e., $\mathsf{C}^i \cap \mathsf{C}^j = \emptyset$, $i \neq j$. This can be solved by simply removing time-overlapping tracks within clusters. The tracks are joined using single-linkage clustering. The benefit of clustering is data reduction and automatic loop closure detection, since these are defined by clusters containing more than one track.

Results

Two feature tracks which stems from the same physical feature are correctly clustered.



The first two tracks represent the same physical feature while the last can be viewed as an outlier.



Interesting extensions to this work could be to consider constrained clustering or an online clustering algorithm.