SVO: Fast Semi-Direct Monocular Visual Odometry

SVO (Semi-direct Visual Odometry) is a **semi-direct monocular** visual odometry algorithm. The algorithm operates directly on pixel intensities, which results in subpixel precision at **high frame-rates**.

1. System overview

The algorithm uses two parallel threads, one for estimating the camera **motion** and a second one for **mapping** as the environment is being explored (Fig 1.). This separation allows fast and constant-time tracking in one thread, while the second thread extends the map, decoupled from hard real-time constraints.

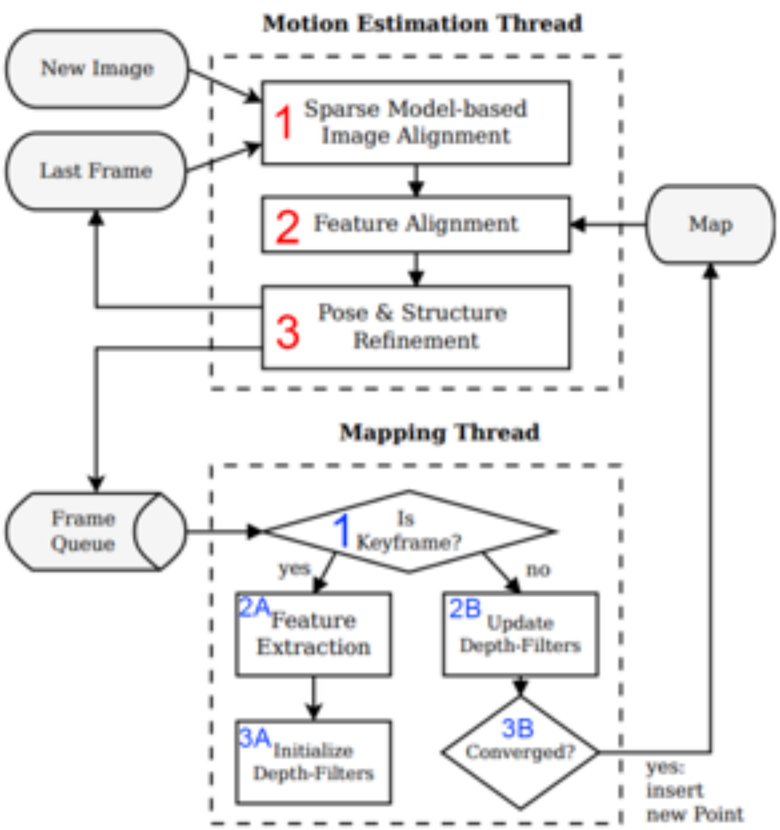

Fig. 1: Tracking and mapping pipeline

2. Motion estimation

1. Sparse model-based image alignment

In this step a rough estimation of the 6 DoF camera pose relative to the previous camera pose is determined. First the photometric difference between the new image and the last frame is calculated to identify features (fig 3.1). Then the photometric difference is minimized by changing the relative pose (fig 3.2). Besides the pose, the 3D coordinates of the features are estimated.

2. Feature Alignment *

First keyframes that contain one or more features from the new image that were identified in step 1, are collected from the map. These keyframes are then used to optimise the 3D coordinates and, thus, the feature coordinates on the new image.

3. Pose & Structure Refinement *

In this step the feature coordinates on the new image that were introduced in the previous step are used to further refine the pose.

* These steps are skipped when there are no keyframes available in the map.

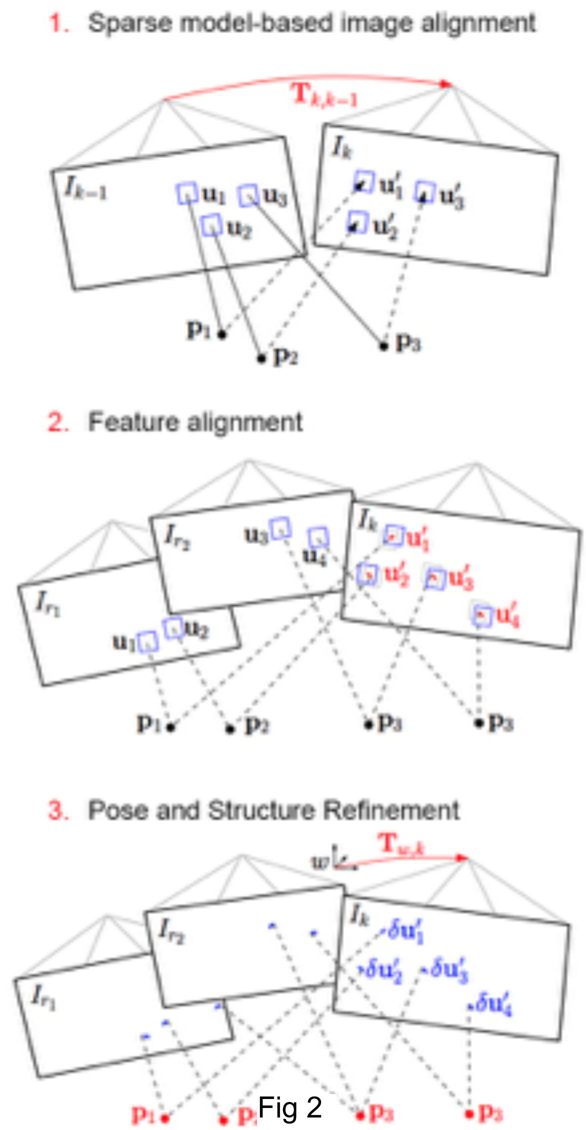

Fig 2

Fig 3.1

Fig 3.2

3. Mapping

1. **Is keyframe?**

A keyframe is selected if the Euclidean pose distance of the new frame relative to all keyframes exceeds 12% of the average scene depth. When a new keyframe is inserted in the map, the keyframe farthest apart from the current position of the camera is removed.

2. A. **Feature extraction**

The keyframe will in this step be divided into cells of a fixed size and for each cell the FAST corner with the highest Shi-Tomasi score is registered unless there is already a 2D-to-3D correspondence present in the map. This results in evenly distributed features in the image.

3. A. **Initialize Depth-Filters**

A new depth-filter is initialized for each FAST corner registered in the previous step.

2. B. **Update Depth-Filters**

In this step the frame, which is not a keyframe, is used to update every depth filter associated with the current keyframe using the Bayesian framework.

3. B. **Converged?**

When the variance of the distribution in a depth filter becomes small enough, the depth-estimate is converted to a 3D point. The point is inserted in the map and immediately used for motion estimation.

SVO 2.0

SVO 2.0 is the successor of SVO. The SVO 2.0 algorithm supports multiple cameras, edge tracking, motion priors and very large field of view cameras, such as fisheye and catadioptric ones.