Robust Selective Stereo SLAM without Loop Closure and Bundle Adjustment

Fabio Bellavia†, Marco Fanfani‡, Fabio Pazzaglia† and Carlo Colombo‡, University of Florence, Italy
†{bellavia.fabio, fabio.pazzaglia}@gmail.com, ‡{marco.fanfani, carlo.colombo}@unifi.it

COMPUTATIONAL VISION GROUP
Università degli Studi di Firenze

SSLAM

Our *Selective SLAM* (SSLAM) framework combines a robust loop chain matching scheme for tracking keypoints with an effective frame selection strategy.

- SSLAM relies on the observation that the pose estimation errors propagate from the uncertainty of the 3D points, higher for distant points corresponding to matches with low temporal flow disparity.
- SSLAM does not require any loop closure or bundle adjustment.
- SSLAM alternates between keypoint matching between consecutive SLAM frames and the estimation of the relative camera poses.

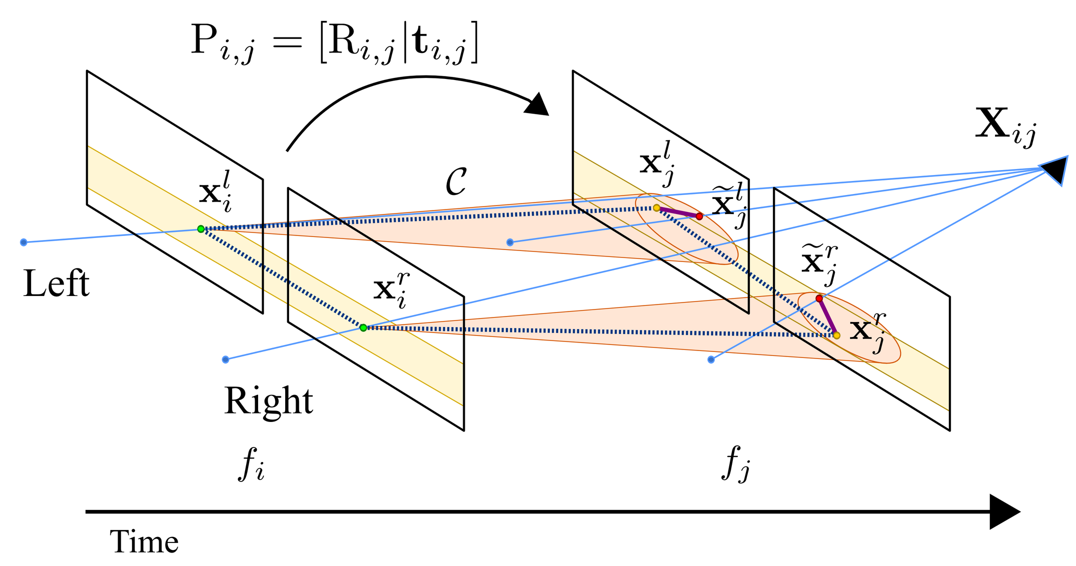
Keypoints detection and matching

The HarrisZ detector and the sGLOH descriptor are used to extract and match respectively stable corner features. Spatial and temporal constrains have been imposed to refine the candidates matches.

- Keypoint matches between the frame f_i and f_j must satisfy the spatial constraint imposed by the epipolar rectification (yellow band) as well as the temporal flow motion restriction (orange cone).
- Furthermore, the four matching points must form the loop chain $\mathcal{C}$ (dotted line)

$$\mathcal{C} = \left((\mathbf{x}_i^l, \mathbf{x}_i^r), (\mathbf{x}_i^l, \mathbf{x}_j^l), (\mathbf{x}_j^l, \mathbf{x}_j^r), (\mathbf{x}_i^r, \mathbf{x}_j^r)\right)$$

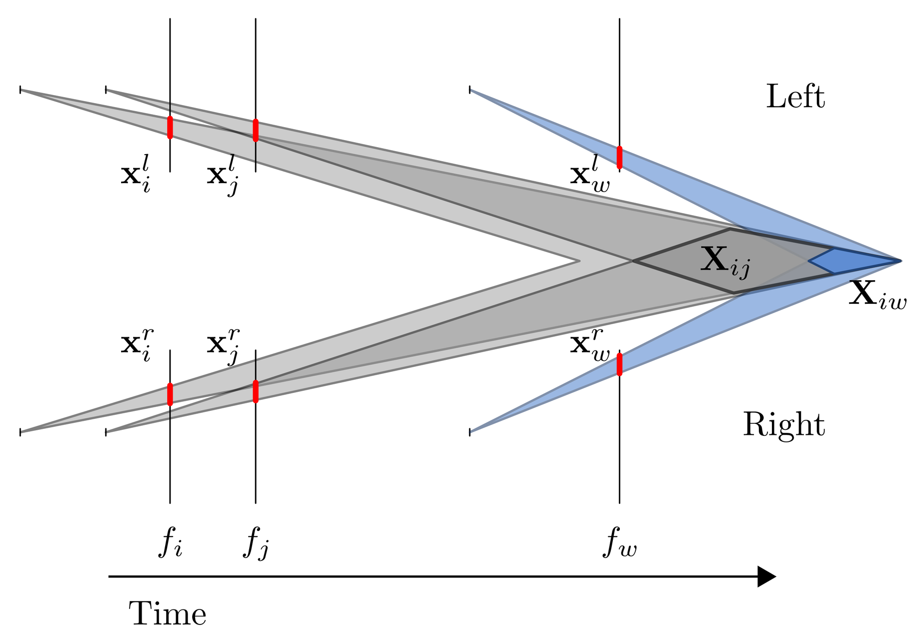
- Finally, candidate matches are refined by running four distinct RANSAC and a subset of chain matches $C_{i,j} \subseteq \{\mathcal{C}\}$ is selected.



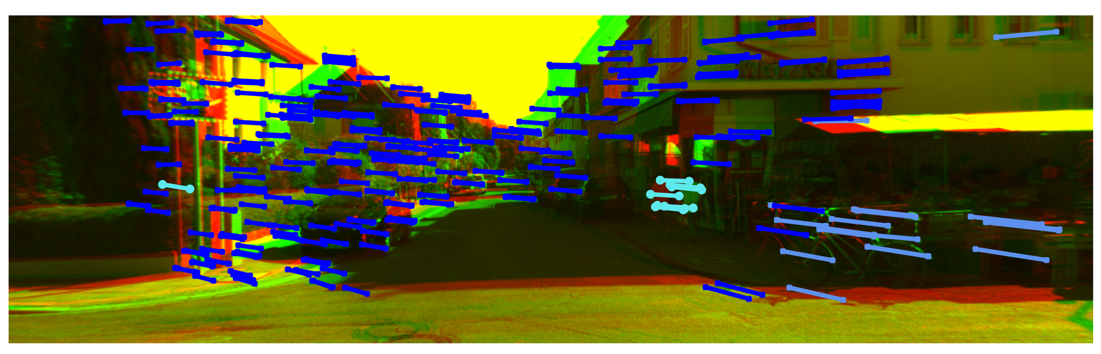
- In the ideal case, points $\mathbf{x}_j^l$, $\mathbf{x}_j^r$ in frame f_j must coincide with the projections $\widetilde{\mathbf{x}}_j^l$, $\widetilde{\mathbf{x}}_j^r$ of points $\mathbf{x}_i^l$, $\mathbf{x}_i^r$ in f_i obtained by triangulation of $\mathbf{X}_{i,j}$ in order for the chain $\mathcal{C}$ to be consistent with the pose $\mathrm{P}_{i,j}$. Due to data noise, in the real case the distances $\| \widetilde{\mathbf{x}}_j^l - \mathbf{x}_j^l \|$ and $\| \widetilde{\mathbf{x}}_j^r - \mathbf{x}_j^r \|$ are minimized by RANSAC on matches $C_{i,j}$.

Pose estimation constrained by temporal flow

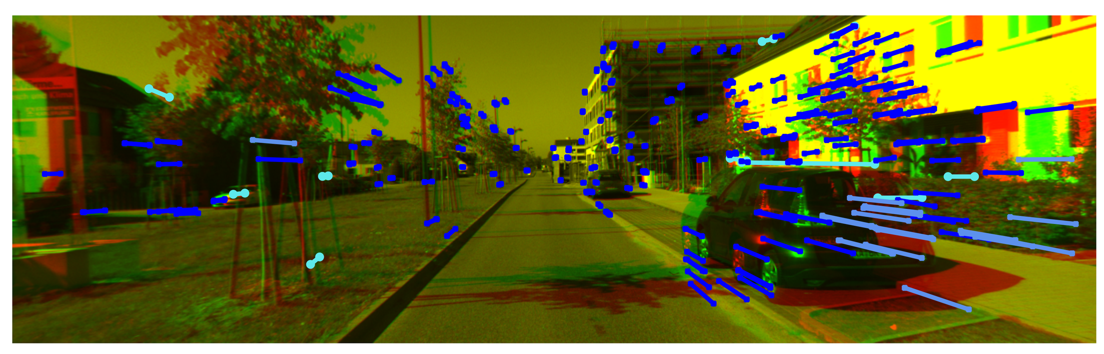
- The uncertainty of matches in the image planes is lower bounded by the image resolution (red) and is propagated to the 3D points. For the same resolution limits, 3D point locations can assume an higher range $\mathbf{X}_{i,j}$ of values (dark gray quadrilateral) when estimated by close frames f_i and f_j, while the possible locations $\mathbf{X}_{i,w}$ are more circumscribed (blue quadrilateral) in the case of distant frames f_i and f_w.



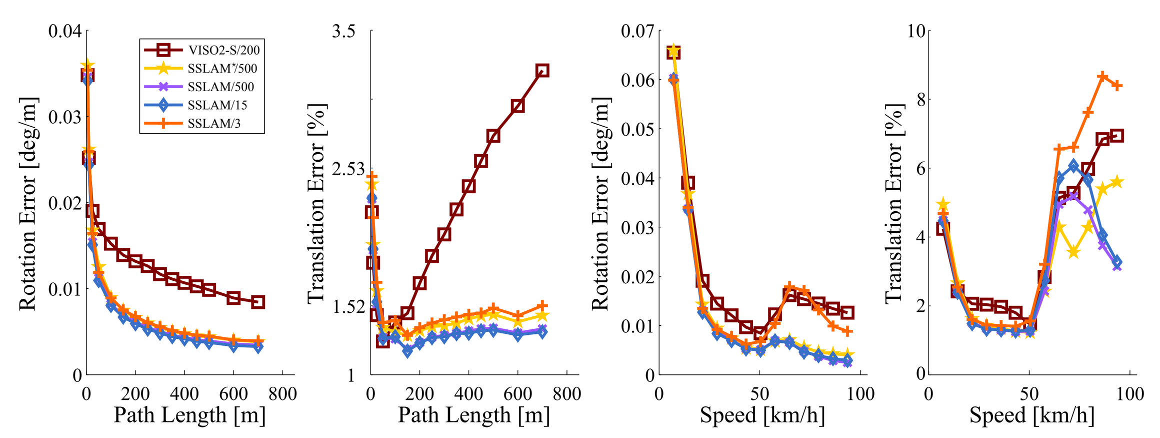
- An adaptive threshold δ_m is defined on the previous observation to discard bad frames containing slow motion information. The frame f_j is discarded and the next frame f_{j+1} is tested if it contains a high percentage of fixed point matches, i.e. for which the distance on the image is below a threshold δ_f.
- Examples of good frames are presented in the figures below, where good fixed and unfixed matches are shown in blue and light violet, respectively, while wrong correspondences are reported in cyan.



- Finally, a pose smoothing constraint between frames is introduced, so that the current relative pose estimation $\mathrm{P}_{i,j}$ cannot abruptly vary from the previous one. This is achieved by imposing that the relative rotation around the origin between the two incremental rotations $\mathrm{R}_{z,i}$ and $\mathrm{R}_{i,j}$ is bounded, as well as for the corresponding translation directions $\mathbf{t}_{z,i}$ and $\mathbf{t}_{i,j}$.
- This last constraint can resolve issues in the case of no camera movement or when moving objects crossing the camera path cover the scene.

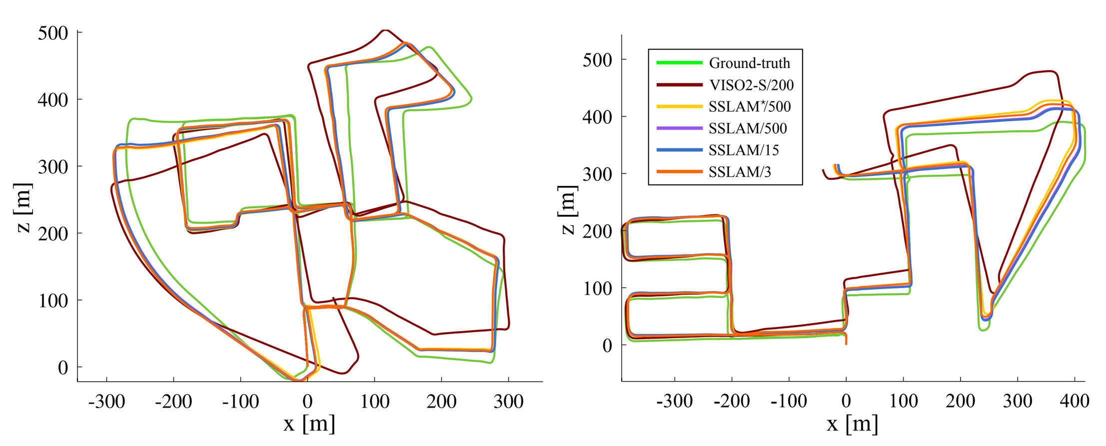


Results



Different versions of SSLAM have been compared with VISO2-S [1], which implements a similar loop chain matching scheme and RANSAC pose estimation on the first 11 sequences of the KITTI vision benchmark suite [2], which provide ground-truth data. SSLAM* refers to SSLAM without the adaptive frame discard detection, the number of RANSAC iterations are also reported.

- Both the different versions of the proposed method provide results better than VISO2-S according to the average translation and rotation errors for increasing path length and speed.
- The chain loop matching scheme together with the chosen keypoint detector and descriptor is robust even for long paths, without requiring bundle adjustment or loop closure detection.
- Dropping low temporal flow disparity frames in SSLAM improves on the standard pose estimation used in SSLAM*, allowing the tracking of longer paths.
- Results for SSLAM/15 and SSLAM/500 are equivalent, while SSLAM/3 obtains inferior results but similar to those obtained by SSLAM*/500, giving an evidence of the robustness of the methods.
- Estimated paths of the different SLAM method on the KITTI dataset show that both SSLAM and SSLAM* paths are closer to the ground-truth than VISO2-S.



Funding

This work has been carried out during the THESAURUS project, founded by Regione Toscana (Italy) in the framework of the "FAS" program 2007-2013.

References

1. A. Geiger et al., "StereoScan: Dense 3D reconstruction in real-time", IEEE Intelligent Vehicles Symposium (2011).
2. A. Geiger et al., "Are we ready for autonomous driving? The KITTI vision benchmark suite", Computer Vision and Pattern Recognition (2012).