Stereo Visual Odometry by Combining Points and Lines

Gomez-Ojeda, Ruben, and Gonzalez-Jimenez, Javier. - University of Malaga
{rubengooj}@gmail.com

ICVSS 2016
Sicily ~ 17 - 23 July
International Computer Vision Summer School

Abstract

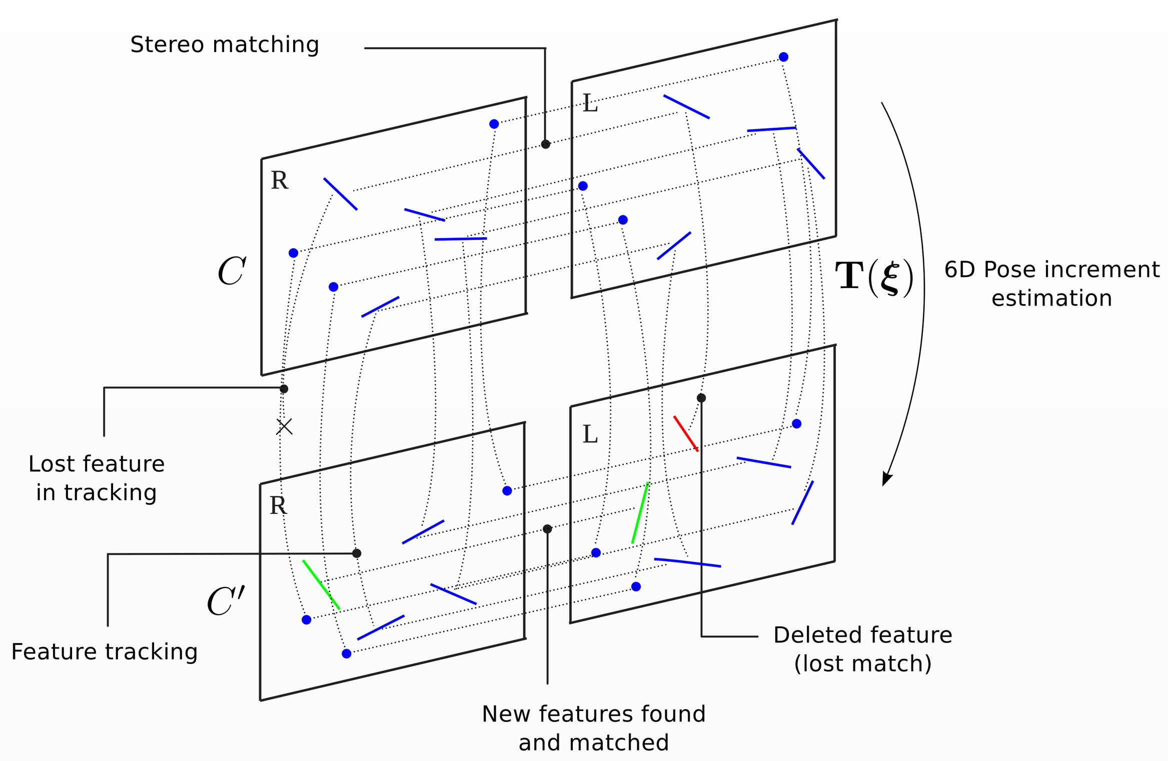
Most approaches to stereo visual odometry reconstruct the motion based on the tracking of point features along a sequence of images. However, in low-textured scenes it is often difficult to encounter a large set of point features, or it may happen that they are not well distributed over the image, so that the behavior of these algorithms deteriorates. This paper proposes a probabilistic approach to stereo visual odometry based on the combination of both point and line segment that works robustly in a wide variety of scenarios. The camera motion is recovered through non-linear minimization of the projection errors of both point and line segment features. The method, of course, is computationally more expensive that using only one type of feature, but still can run in real-time on a standard computer and provides interesting advantages, including a straightforward integration into any probabilistic framework commonly employed in mobile robotics.

Motivation

Typical outdoor scenario, where both types of features are abundant.

Illustrative example of a scene without point features.

System Overview



Scheme of the feature tracking process

Combined Stereo Visual Odometry

The problem we face is that of estimating the optimal $\boldsymbol{\xi} \in \mathfrak{se}(3)$ that minimizes the projection error for points and line segments. The point projection error $\Delta\mathbf{p}_i(\boldsymbol{\xi})$ is given by:

$$\Delta\mathbf{p}_i(\boldsymbol{\xi}) = \hat{\mathbf{p}}_i(\boldsymbol{\xi}) - \mathbf{p}'_i \tag{1}$$

with $\mathbf{p}'_i$ being the i-th detected point in the second frame, and $\hat{\mathbf{p}}_i(\boldsymbol{\xi})$ the projected point from the first frame to the second one. The line projection error is defined as:

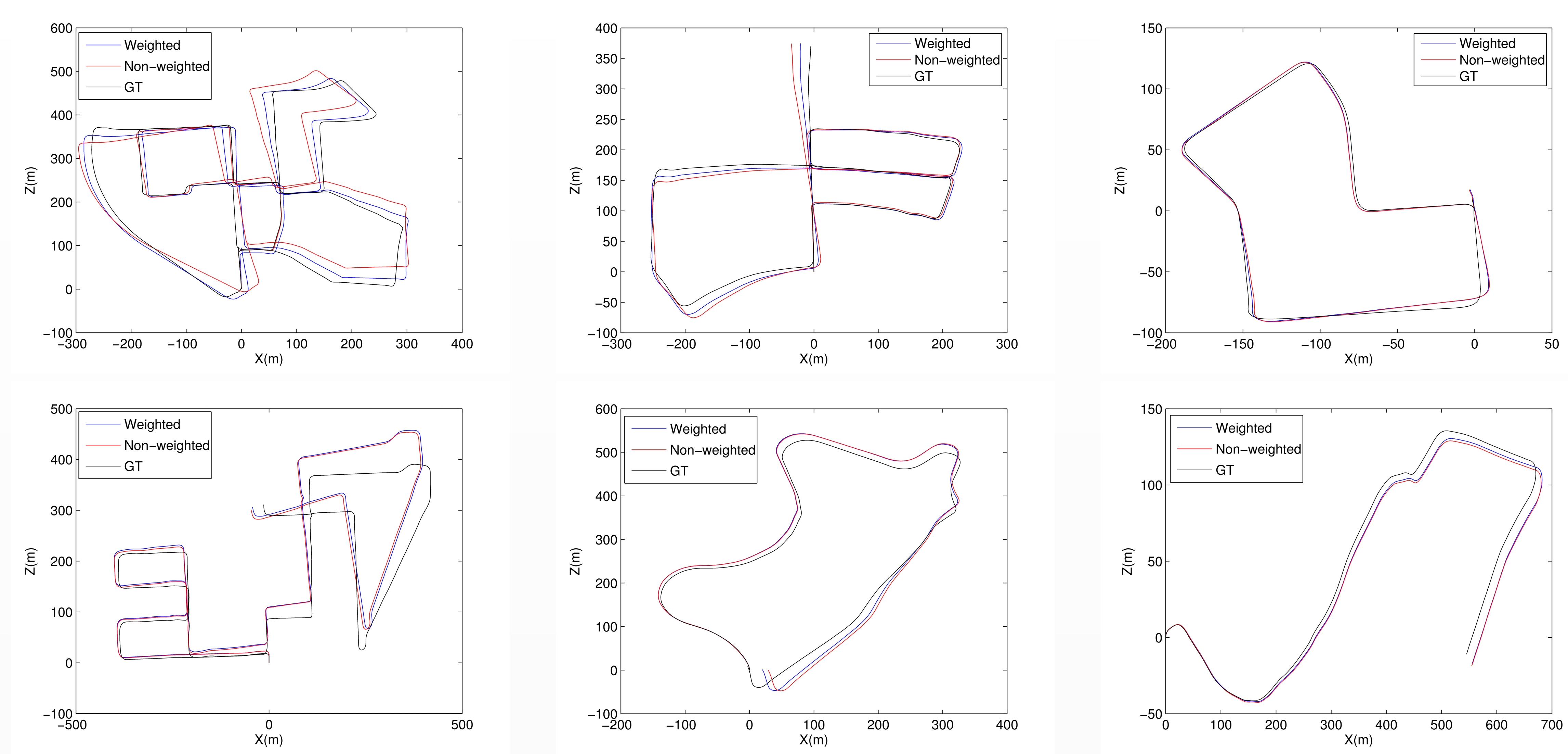
$$\Delta\mathbf{l}_j(\boldsymbol{\xi}) = \left[\mathbf{l}'^{\top}_j \cdot [\hat{\mathbf{p}}_j(\boldsymbol{\xi}) \;\; \hat{\mathbf{q}}_j(\boldsymbol{\xi})]\right]^{\top} \tag{2}$$

where $\hat{\mathbf{p}}_j(\boldsymbol{\xi})$ and $\hat{\mathbf{q}}_j(\boldsymbol{\xi})$ refer to the projected endpoints, and $\mathbf{l}'_j$ is the j-th infinite line detected in the second frame. The optimal pose increment $\mathbf{T}(\boldsymbol{\xi}^*)$ is computed through iterative Gauss-Newton minimization of the reprojection errors, given by the following non-linear least-squares estimator:

$$\boldsymbol{\xi}^* = \underset{\boldsymbol{\xi}}{\text{argmin}} \left\{ \sum_i^{N_p} \Delta\mathbf{p}_i(\boldsymbol{\xi})^{\top} \boldsymbol{\Sigma}^{-1}_{\Delta\mathbf{p}_i} \Delta\mathbf{p}_i(\boldsymbol{\xi}) + \sum_j^{N_l} \Delta\mathbf{l}_j(\boldsymbol{\xi})^{\top} \boldsymbol{\Sigma}^{-1}_{\Delta\mathbf{l}_j} \Delta\mathbf{l}_j(\boldsymbol{\xi})^{\top} \right\} \tag{3}$$

where N_p and N_l corresponds to the number of point and line correspondences respectively, and the matrices $\boldsymbol{\Sigma}^{-1}_{\Delta\mathbf{p}_i}$ and $\boldsymbol{\Sigma}^{-1}_{\Delta\mathbf{l}_i}$ are the 2×2 inverse of the covariance matrices for each type of feature.

Experimental Validation



Results in several sequences from the KITTI dataset [1].

References

[1] A. Geiger, P. Lenz, R. Urtasun, Are we ready for autonomous driving? the kitti vision benchmark suite, in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012