

## 1. Spatio-Temporal and Geometry Constrained Network for Automobile Visual Odometry

**Accession number:** 20203309040204

**Authors:** Liu, Hong (1); Wei, Peng (1); Huang, Weibo (1); Hua, Guoliang (1); Meng, Fanyang (2)

**Author affiliation:** (1) Peking University, Key Laboratory of Machine Perception, Shenzhen Graduate School, China; (2) Peng Cheng Laboratory, Shenzhen, China

**Source title:** ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings

**Abbreviated source title:** ICASSP IEEE Int Conf Acoust Speech Signal Process Proc

**Volume:** 2020-May

**Part number:** 1 of 1

**Issue title:** 2020 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2020 - Proceedings

**Issue date:** May 2020

**Publication year:** 2020

**Pages:** 2747-2751

**Article number:** 9052903

**Language:** English

**ISSN:** 15206149

**CODEN:** IPRODJ

**ISBN-13:** 9781509066315

**Document type:** Conference article (CA)

**Conference name:** 2020 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2020

**Conference date:** May 4, 2020 - May 8, 2020

**Conference location:** Barcelona, Spain

**Conference code:** 161907

**Sponsor:** The Institute of Electrical and Electronics Engineers, Signal Processing Society

**Publisher:** Institute of Electrical and Electronics Engineers Inc., United States

**Abstract:** Visual odometry (VO) is an essence of vision-based localization and mapping system where existing learning-based approaches utilize CNN and RNN to model camera motion and gain promising results. However, these methods lack full use of the relationship between spatial characteristics and temporal clues, as well as geometry constraints in VO. To overcome these deficiencies, an end-to-end framework that leverages spatio-temporal relevance and geometrical knowledge is proposed. In particular, a spatial response module (SRM) is designed to extract the visual motion features by emphasizing the most interconnected regions while suppressing the irrelevant areas. A module named temporal response module (TRM) is used to regress the camera motion via adopting the optimal motion features. Moreover, a geometry constrained (GC) loss that minimizes the estimated inter-frame pose errors and the accumulated pose errors within a local period is introduced. Actually, the GC loss utilizes adaptive learnable balance factors for balancing losses. Experimental results on KITTI and Malaga datasets demonstrate that the proposed model outperforms state-of-the-art monocular methods. © 2020 IEEE.

**Number of references:** 27

**DOI:** 10.1109/ICASSP40776.2020.9052903

**Funding Details:** Number: 2018YFB1308602, Acronym: -, Sponsor: -; Number: 61673030,61906103,U1613209, Acronym: NSFC, Sponsor: National Natural Science Foundation of China; Number: -, Acronym: NSFC, Sponsor: National Natural Science Foundation of China;

**Funding text:** This work is supported by National Natural Science Foundation of China (NSFC No. U1613209, 61673030, 61906103), National Key RD Program from the Ministry of Science and Technology of China (Grant No.2018YFB1308602).

**Compendex references:** YES

**Database:** Compendex

Compilation and indexing terms, Copyright 2020 Elsevier Inc.

**Data Provider:** Engineering Village