SuA IV : Vision V

Session Chair : Qieshi Zhang, Chengzhi Hu
Tencent ID : 234 889 136 : July 18, 16:15 - 18:00, Sunday
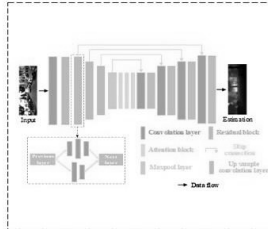
- 16:15~16:30

Attention Mechanism-based Monocular Depth Estimation and Visual Odometry

Qieshi Zhang, Dian Lin, Ziliang Ren, Yuhang Kang, Fuxiang Wu, Jun Cheng
Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China

- Propose a depth estimation network encoder based on attention mechanism for predicting the spatial features of the adjacent frames.
- Improve the depthwise separable convolution to replace the convolution layers in the main architecture for improving efficiency.
- Validate the performance of our algorithm on a public dataset and real-world experiments with our mobile robot, including outdoor and indoor with ground truth.

- 16:30~16:45

Skeleton-based Action Recognition with Multi-scale Spatial-temporal Convolutional Neural Network

Qin Cheng, Ziliang Ren, Jun Cheng, Qieshi Zhang, Hao Yan and Jianming Liu
Guilin University of Electronic Technology, Guilin, China

- Action recognition based on skeleton data needs comprehensive spatial-temporal features.
- A novel MSST-Module is established to capture robust spatial-temporal features.
- The MSST-Net constructed by MSST-Module achieved remarkable performance on two large dataset.

Structure of MSST-Net

- 16:45~17:00

Bidirectional Weighted Loss with Feature Perception for Self-supervised Learning of Consistent Depth-pose

Fei Wang, Jun Cheng and Penglei Liu
CAS Key laboratory of Human-machine Intelligence-Synergy Systems, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China

- Take full advantage of limited data by using the bidirectional photometric loss.
- Deal with moving objects and occlusions by reweighting the bidirectional photometric loss.
- Improve the robustness for textureless regions by employing the bidirectional feature perception loss.
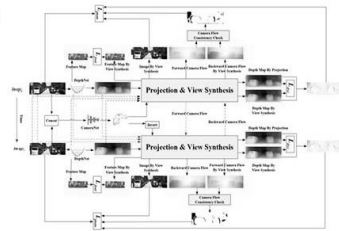- Enforce consistency between depths by employing the bidirectional depth structure consistency loss.
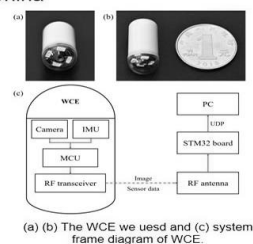
Diagram of the general framework

- 17:00~17:15

Two Stream Dynamic Threshold Network for Weakly-Supervised Temporal Action Localization

Hao Yan, Jun Cheng, Qieshi Zhang, Ziliang Ren, Shijie Sun,Qin Cheng
Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China

- Weakly-supervised methods only require video-level labels to train the models.
- The proposed DH-WTAL features a dynamic attention threshold decision for the attention.
- Our model further adjust the extreme values of the attention mechanism for different videos accordingly.

- 17:15~17:30

Real-time Attitude Tracking of Capsule Endoscope Based on MEMS IMU and Error Analysis

Zhuokang Huang and Chengzhi Hu
Department of Mechanical and Energy Engineering, Southern University of Science and Technology, China

- A method for calculating the attitude of WCE based on MEMS IMU is implemented.
- The accuracy of attitude tracking is measured by a series of experiments.
- The experimental results show that the method meets the requirements in about six minutes.
- Additionally, we propose a method to improve the attitude accuracy by reciprocating the rotational motion.

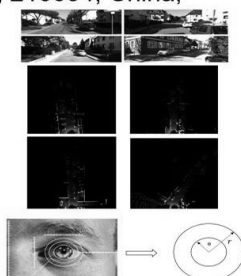(a) (b) The WCE we uesd and (c) system frame diagram of WCE.

- 17:30~17:45

Unseen Object Pose Estimation via Registration

Jun Wu, Yue Wang and Rong Xiong
Control Science and Technology, Zhejiang University, China

- Current object pose estimation methods mostly rely on instance specific features, which limits their ability to generalize to unseen objects.
- 3D geometrical construction is embedded in observation, yet commonly neglected.
- We reconstruct full model and view model from reference and query observation, then seek for probabilistic correspondence between them to solve registration problem.
- Our method achieves comparable performance with SOTA, with accuracy and efficiency

- 17:45~18:00

3D LiDAR Point Cloud Loop Detection Based on Dynamic Object Removal
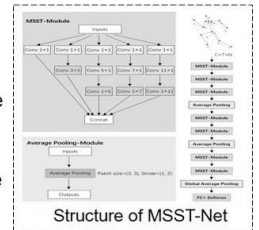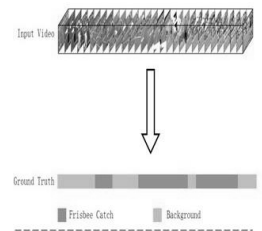
Pan-Yun Ding, Zhi-Hui Wang
School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China;

- 3D object detection model OpenPCDet is employed to detect dynamic objects in the outdoor scene, such as vehicles, pedestrians, etc.
- We use the bounding box detected by the model to perform cube filtering on the original data to remove dynamics objects.
- The processed data is utilized to extract scene descriptors for loop detection. In the road scene, experimental results demonstrate that our approach yields superior performance against the traditional methods.