**CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction**

**Introduction**

* **Volumetric fusion approaches** – Kinect Fusion, dense SLAM methods based on RGB-D

**Limitations:** limited working range poor performance under sunlight

* **B. Monocular SLAM approaches**

These approaches aim at real-time monocular scene reconstruction by estimating the depth map of the current viewpoint through small-baseline stereo matching over pairs of nearby frames. The working assumption is that the camera translates in space over time, so that pairs of consecutive frames can be treated as composing a stereo rig. Stereo matching is usually carried out through color consistency or by relying on key point extraction and matching.

**Limitations:**

Wrong estimation of the absolute scale.

Tracking failures under pure rotational camera motion

* C. Deep learning approaches

**Advantages:**

absolute scale can be learned from examples and thus predicted from a single image without the need of scene-based assumptions or geometric constraints.

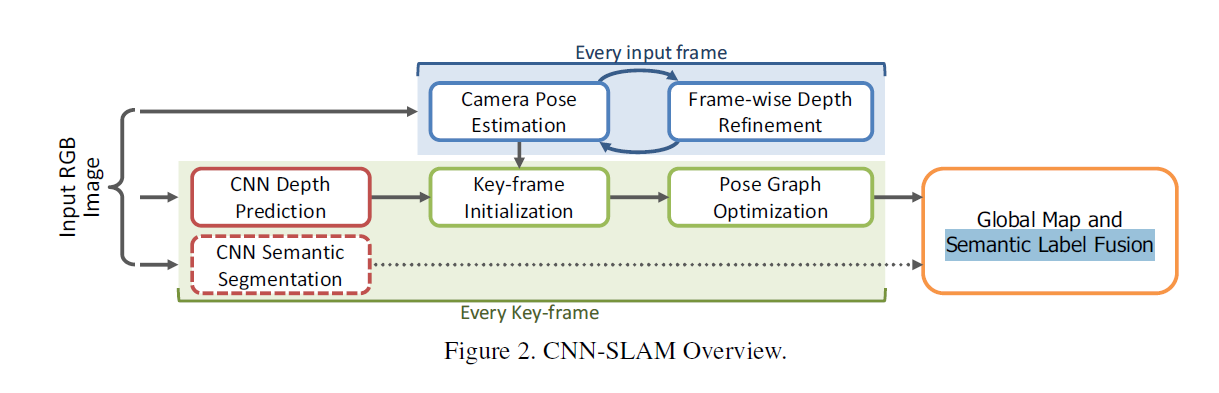
Refined by small-baseline stereo matching

**Limitations:**

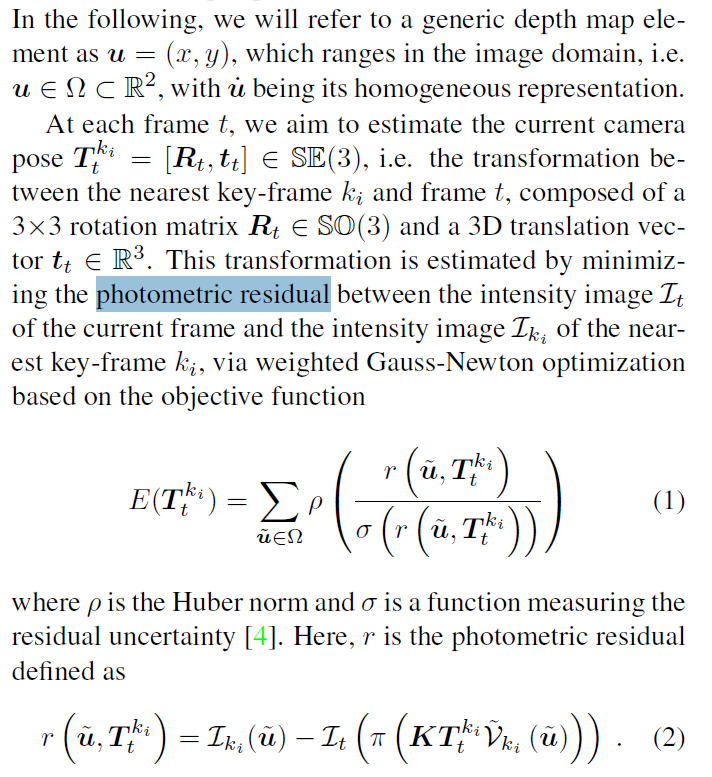
Although globally accurate, depth borders tend to be locally blurred: hence, if such depths are fused together for scene reconstruction as in, the reconstructed scene will overall lack shape details.

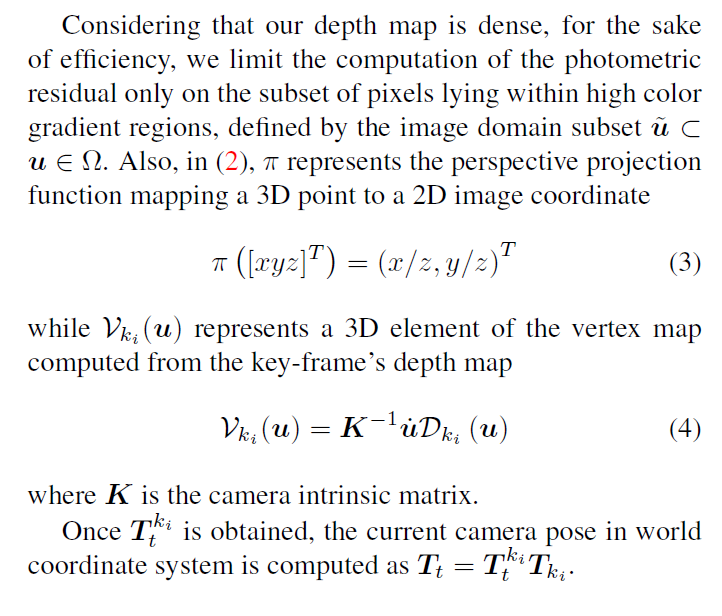
**Related work**

* CNN-SLAM Overview



1. **Camera Pose Estimation**

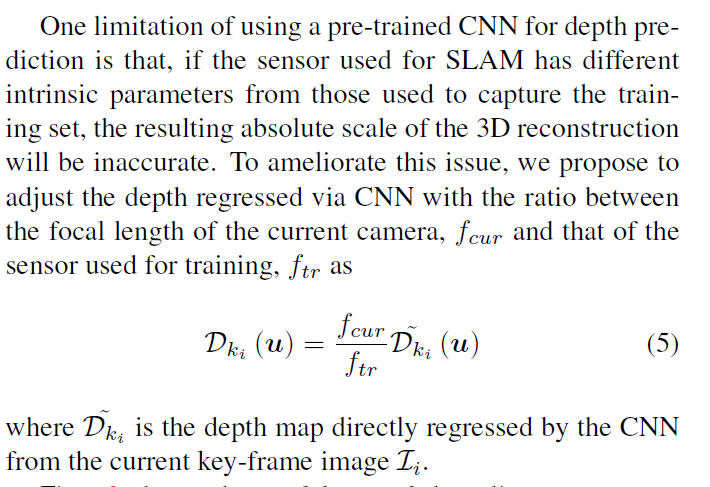


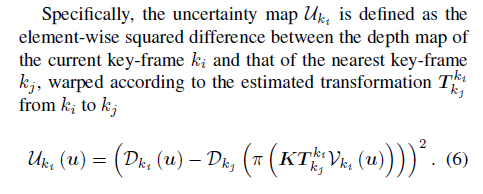


1. **CNN-based Depth Prediction and SEMANTIC Segmentation**

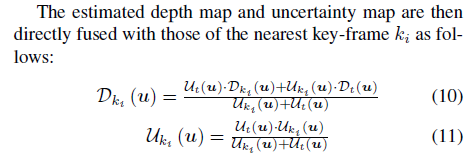
Based on ResNet-50 and initialized with pre-trained weights on ImageNet

1. **Key-frame Creation and Pose Graph Optimization**





1. **Frame-wise Depth Refinement**



1. **Global Model and Semantic Label Fusion**

