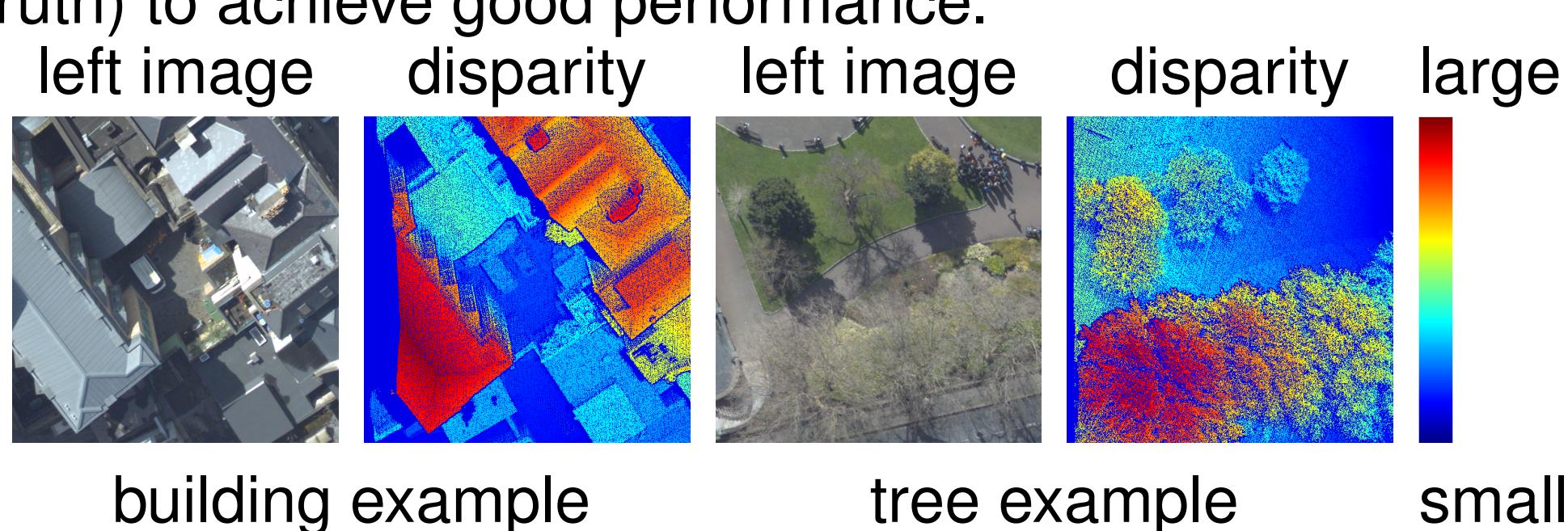


Introduction

Dense matching is a traditional topic in 3D reconstruction, in this work, we focus on binocular epipolar stereo dense matching. To evaluate stereo dense matching methods, many benchmark datasets have been released, especially in computer vision community, for example Middlebury, ETH3D, KITTI and so on.

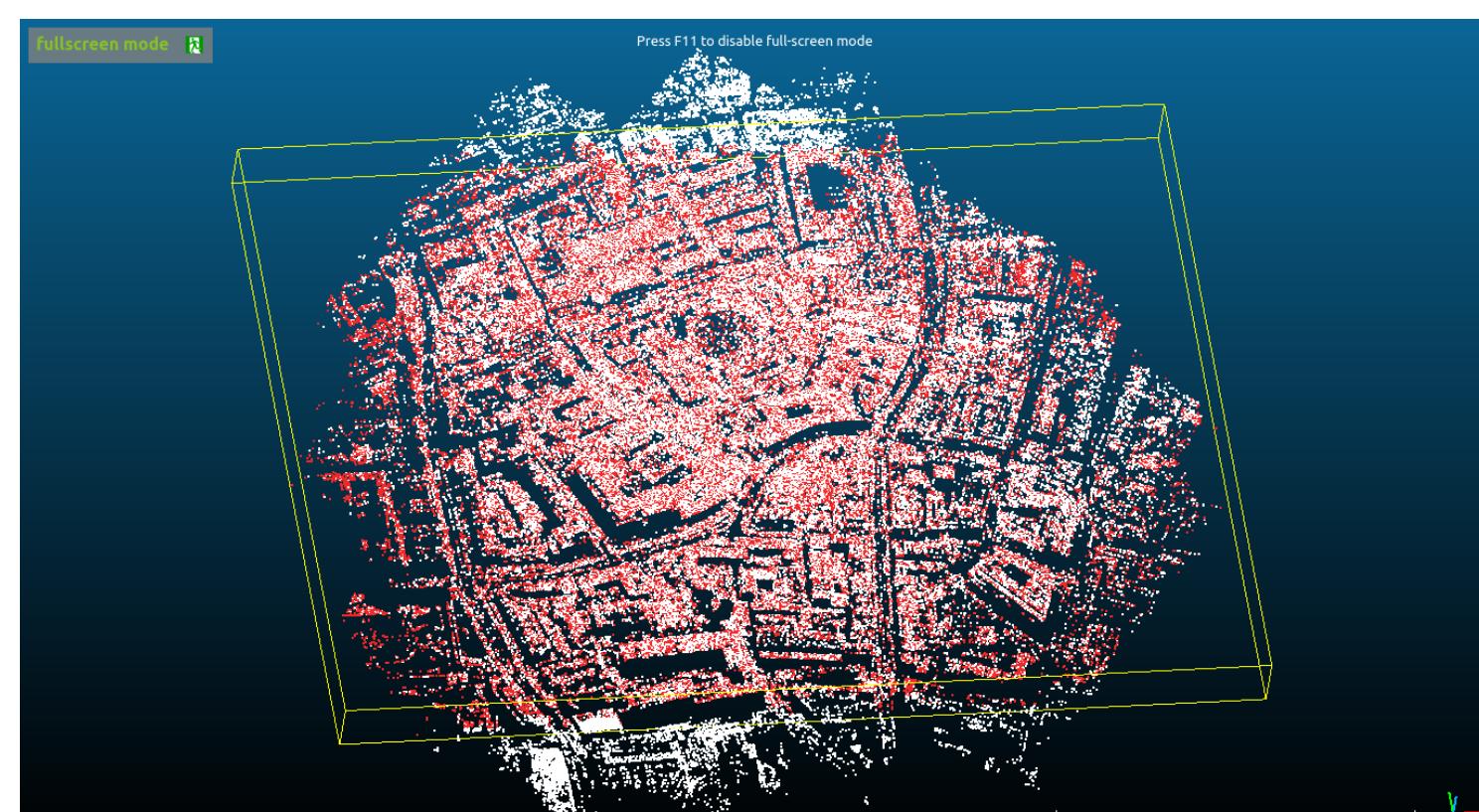
But it is difficult to find good quality aerial stereo dataset for training. While machine learning, and especially deep learning based dense matching methods outperform traditional ones, they usually need a lot of training data (ground truth) to achieve good performance.



ICP based Image-LiDAR registration

If the image and LiDAR is not well-registered, we propose to refine the registration in order to improve their consistency.

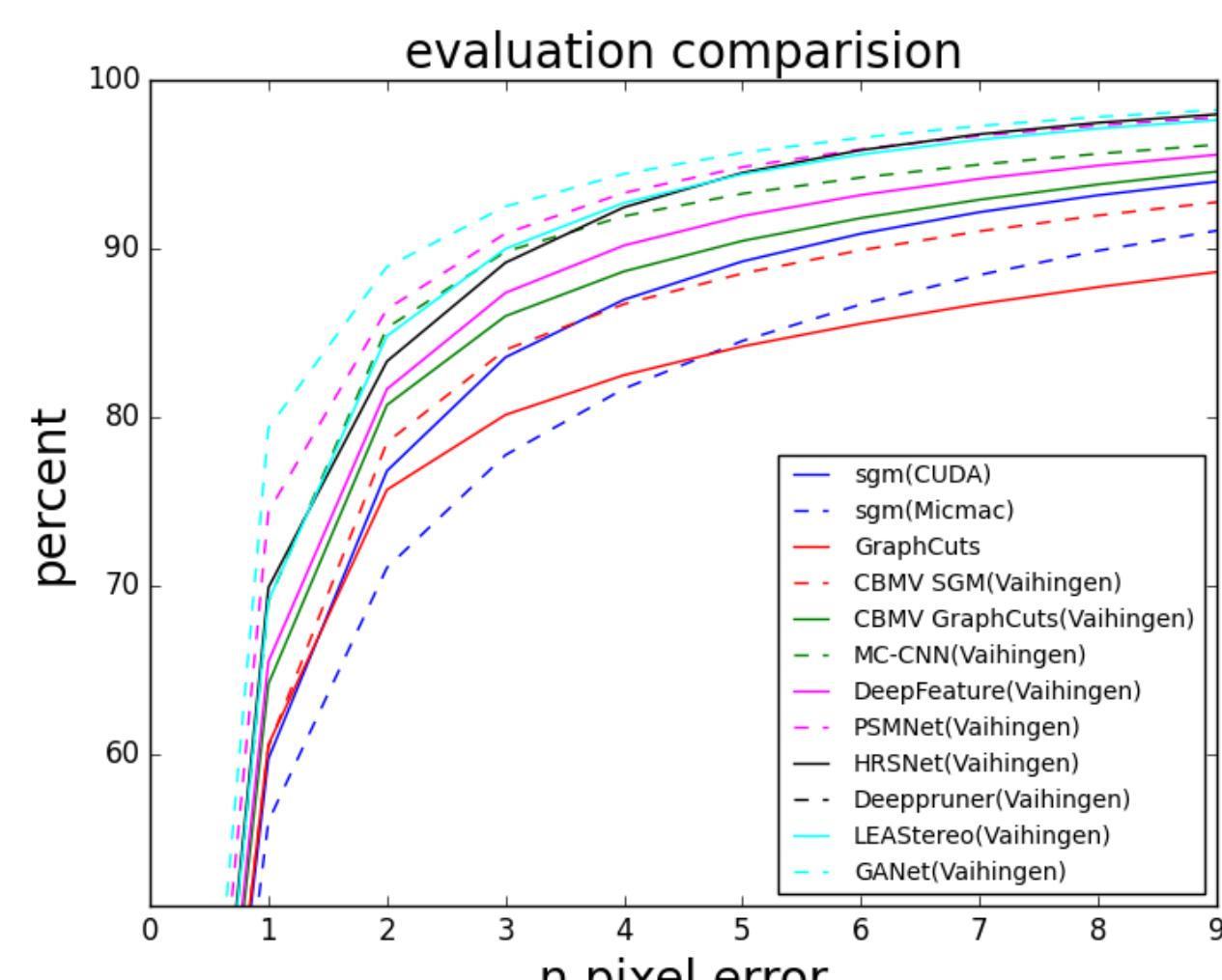
We start by computing the 3D tie points corresponding to the feature points matched during the bundle adjustment. As the initial orientation is close enough to the real position, these tie points lie close to LiDAR points. As the noise in LiDAR is low, so these closest Lidar points are accurate and they can be regarded as a ground control points(GCPs).



GCPs in registration. White points are features points, and red points are the nearest points in LiDAR.

Experiment setting

We extended our previous work [4] by evaluating 11 stereo matching methods, among them 7 additional recent deep learning based methods [1, 3]. Six datasets over 4 different areas were tested.



Tests performed on ISPRS Vahingen.

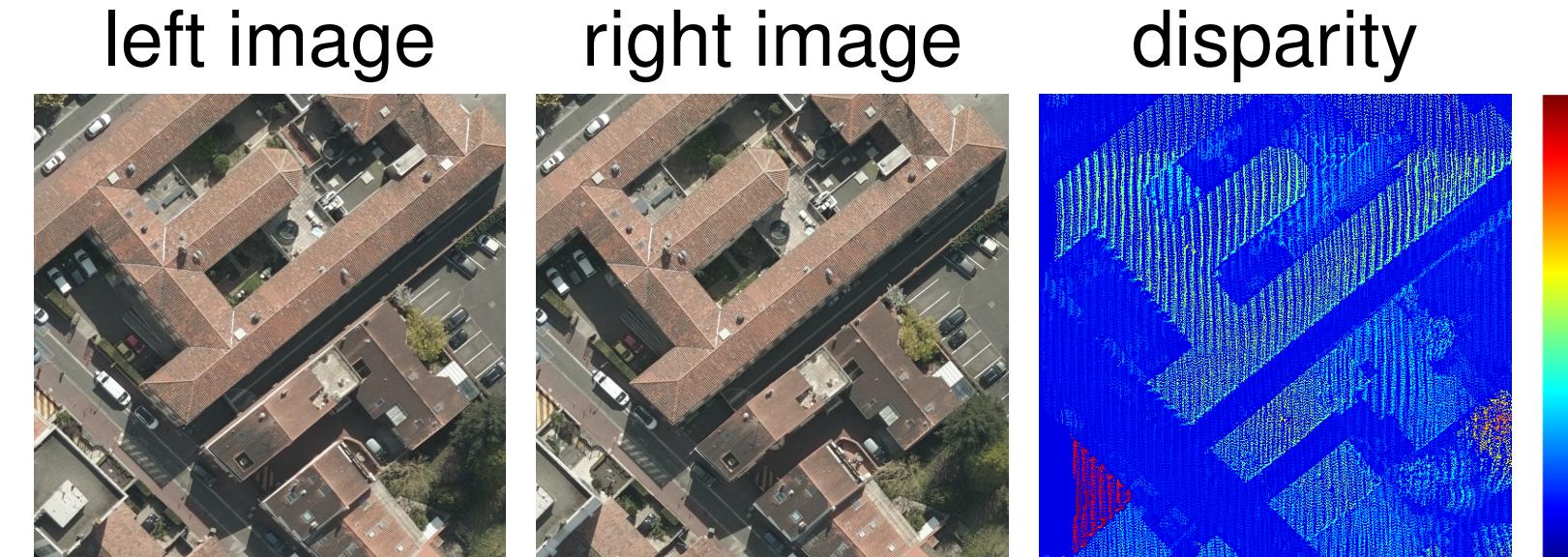
Benchmarking

There are two parts in this work :

1. Generate benchmark from aerial dataset
2. Experiments on SOTA deep learning(DL) methods and dataset shift for DL methods

Work flow of ground truth generation is :

1. Image-LiDAR registration
2. Epipolar image generation
3. Occlusion aware point cloud projection
4. Remove areas with scene changes

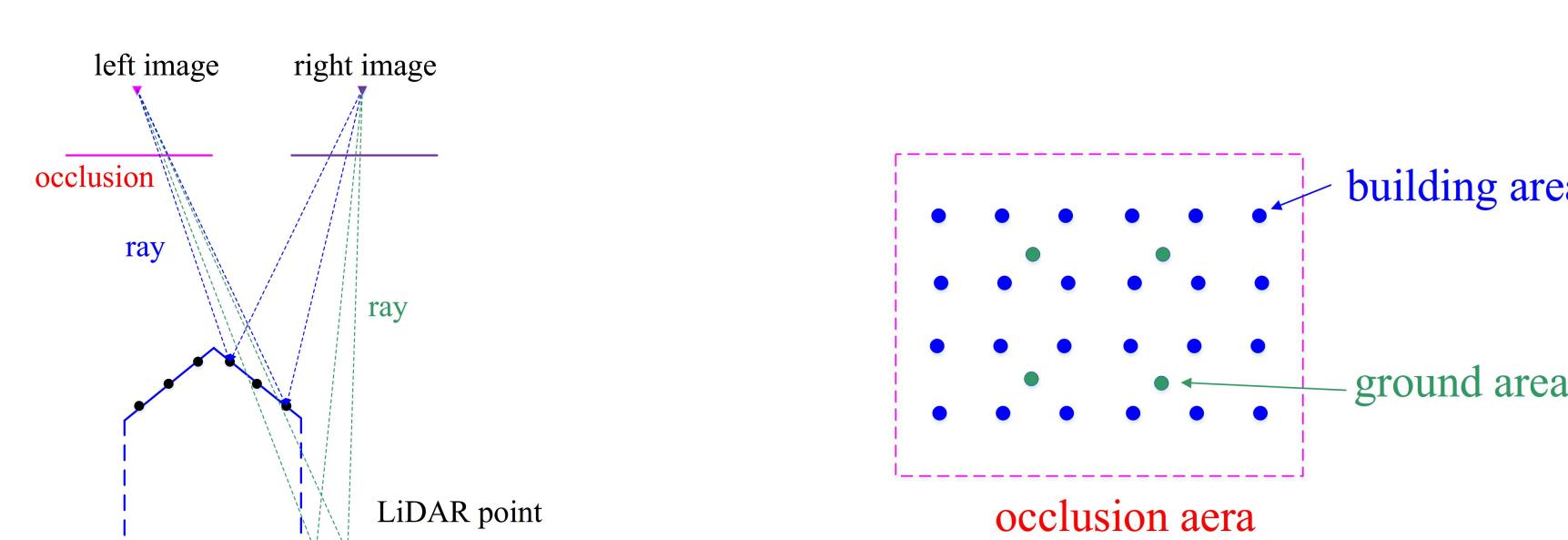


Ground truth generation

Epipolar creation & Occlusion aware projection

In our experiment, only images pairs with an overlap above half the image footprint are considered for epipolar rectification, which was done using a recent module of the MicMac library [2].

The main question of occlusion aware projection is to ensure that a Lidar point is only projected in an (epipolar) image if the corresponding 3D point is actually seen in this image. In the method, we propose a ray tracing based occlusion detection before applying the density based filter.



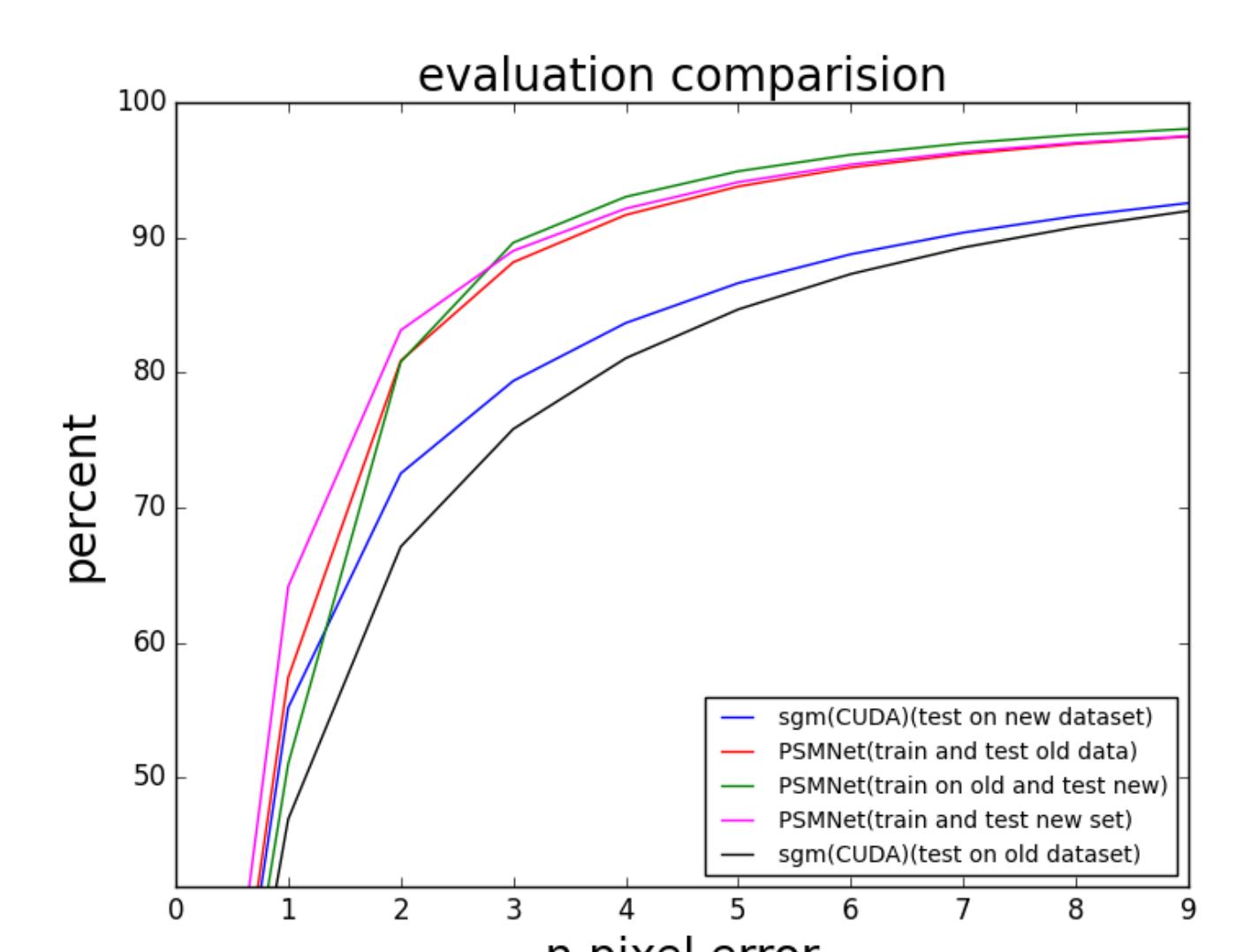
(a)occlusion on roof (b)density is different in occlusion area

Reference

- [1] Jia-Ren Chang and Yong-Sheng Chen. Pyramid stereo matching network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5410–5418, 2018.
- [2] Marc Deseilligny and Ewelina Rupnik. Epipolar rectification of a generic camera. 2021.
- [3] Shivam Duggal, Shenlong Wang, Wei-Chiu Ma, Rui Hu, and Raquel Urtasun. Deeppruner: Learning efficient stereo matching via differentiable patchmatch. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4384–4393, 2019.
- [4] T Wu, B Vallet, M Pierrot-Deseilligny, and E Rupnik. a new stereo dense matching benchmark dataset for deep learning. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43:405–412, 2021.

Remove areas with scene changes

In the experiment, we found that scene change influence performance a lot, if there is no change, the evaluation is good, but if there is an change, it will bring huge errors. First, we use SGM(CUDA) to select high accuracy (with high evaluation score) stereo pairs. Then, we fine-tune PSMNet on these pairs, then use the fine-tuned model on the whole dataset. Finally the stereo pairs with large errors are removed.

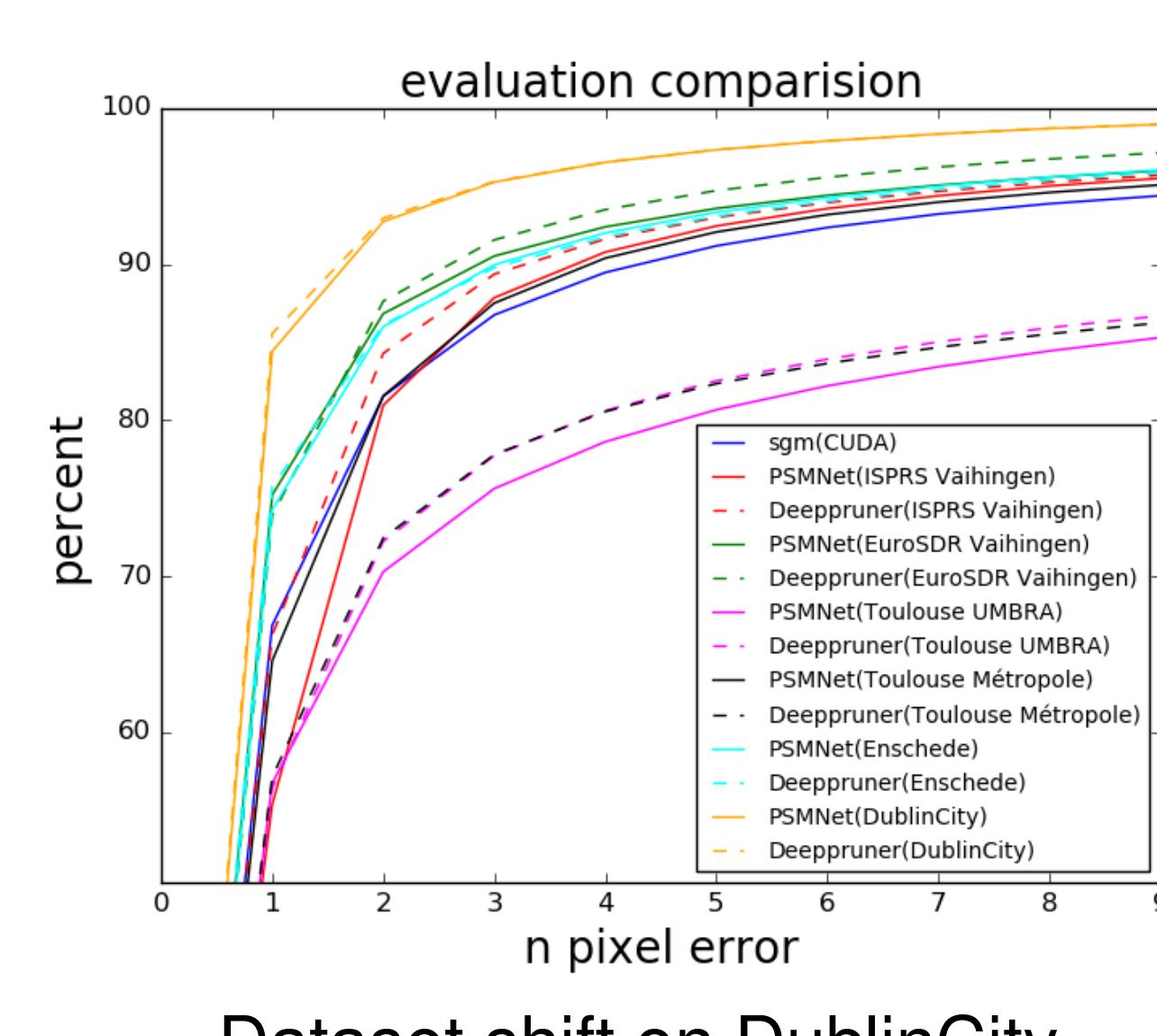


Scene change influence performance.

Experiment and analysis

Dataset shift

In practice, it is difficult/impossible to obtain training dataset for every possible scene. Training a model on one scene and applying it to a different scene is a common scenario. The transferability of the trained model or data-shift are extremely important.



Dataset shift on DublinCity.

Visual assessment

