Enhancing Depth Quality of Stereo Vision using Deep Learning-based Prior Information of the Driving Environment

Vijay John
Research Center for Smart Vehicles
Toyota Technological Institute, Nagoya, Japan

豊田工業大学
TOYOTA TECHNOLOGICAL INSTITUTE

Presentation Outline

- Introduction
- Outline of proposed method
- Experimental results
- Conclusion

Introduction

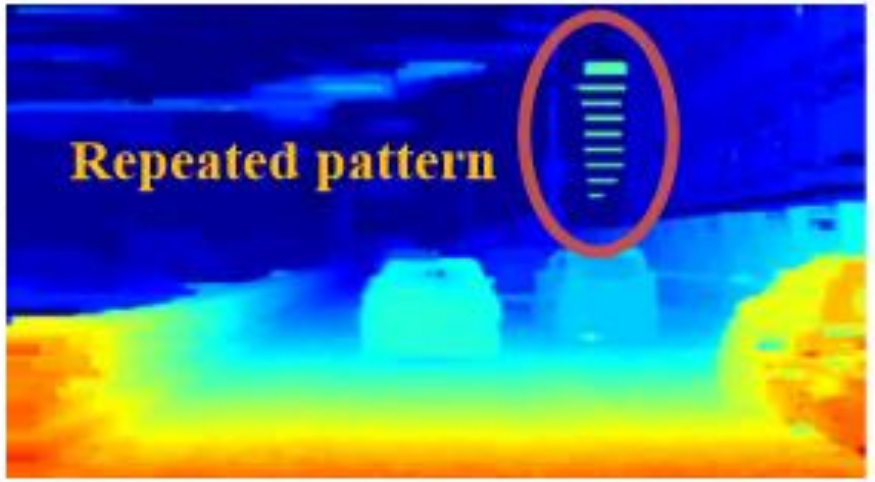
To achieve accurate driving maneuver obtaining dense distance information from the surrounding environment is indispensable.

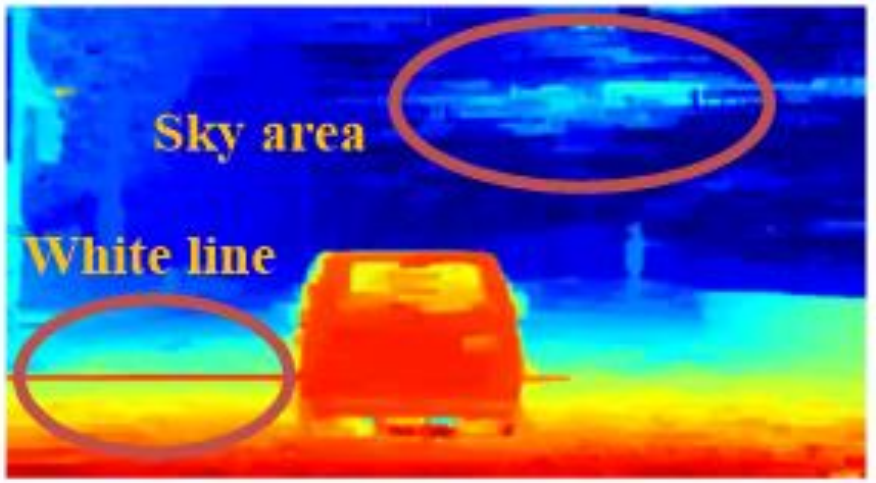
Currently, there are many surrounding environment detection methods using camera, lidar, and radar.

Among these sensors, stereo vision based on binocular camera system can be regarded as one of the leading methods.

Big Challenges. The accuracy of the stereo vision is limited by texture-less regions, such as sky and road areas, and repeated patterns.

Introduction





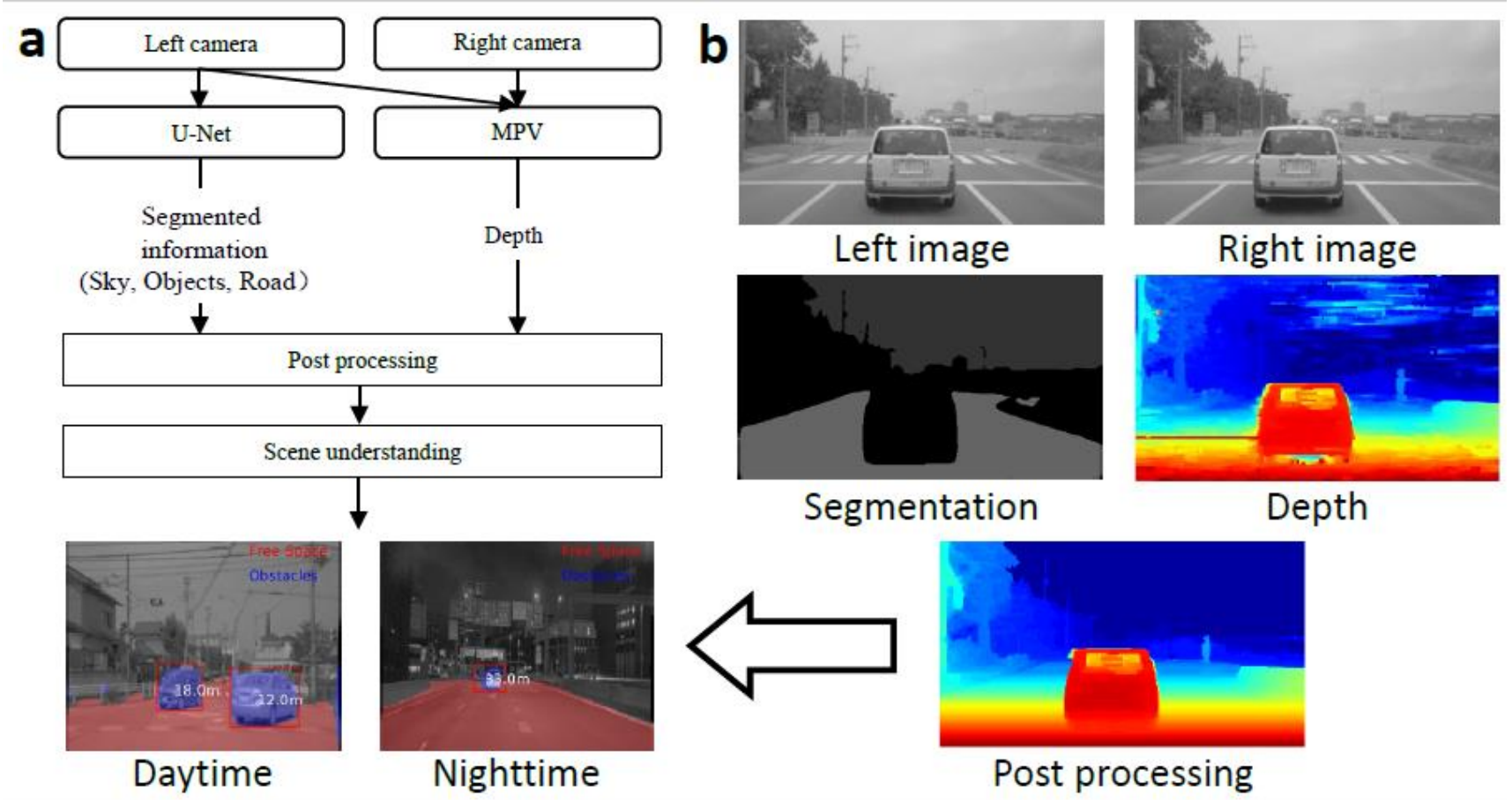
Introduction

We propose to enhance the stereo generated depth by incorporating prior information of the driving environment.

Main contributions.

1. Introduction of adaptive smoothing within the Multi Path Viterbi (MPV) algorithm.
2. Incorporation of deep learning-based semantic segmentation within the MPV algorithm.
3. A novel mathematical optimization model for the post-processing framework which is explainable.

Outline of proposed method



Step 1: Outline of proposed method

- **U-Net Segmentation**
  - We adopt the U-Net to localize the sky and road areas.
  - In the encoder, feature extraction branches, the image features are extracted using multiple 2D convolutional and pooling layers.
  - These features are transferred to the decoder branch using the skip connection.
  - The decoder contains multiple decoding convolutional layers and up-sampling layers.
  - Categorical cross entropy is used as the loss function in the final layer.

Step 2: Outline of proposed method

- **MPV Algorithm**

The MPV algorithm find a disparity map u that minimizes the energy function $E(u)$ as follows:

$$E(u) = \sum_{p} SSIM(p, u) + \sum_{p' \in L_p} \lambda_p \exp(-|P|)|u - u'|$$

The first term is the sum of all pixel matching costs for the disparities of u the structural similarity index (SSIM) cost function.

The second term is the TV constraint modified by the gradient information P of the left image.

It penalizes all the disparity changes between p and p'which has disparity u' and belongs to p's neighbourhood L_p.

λ_p is the parameter to control the smoothness, which is dependent on the segmentation information of left image.

Step 3: Outline of proposed method

- **Post-processing Framework**
- ***Sky area:***
- We assign the disparity values of sky area to be zero.
- Since the sky area is always very far away and the disparity should be zero.

Step 3: Outline of proposed method

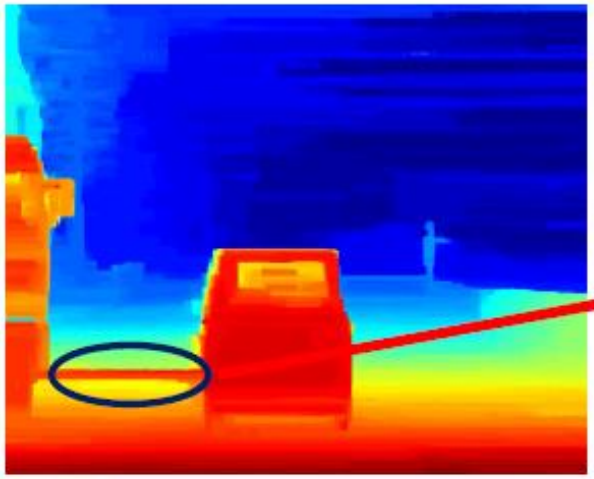
- **Post-processing Framework**
- ***Road area:***
- We assign the disparity values for each horizontal line of the road area using the local and global disparity values of the road area.
- The disparity values for each horizontal line is assigned such that the disparity values are strictly increasing in the vertical direction.

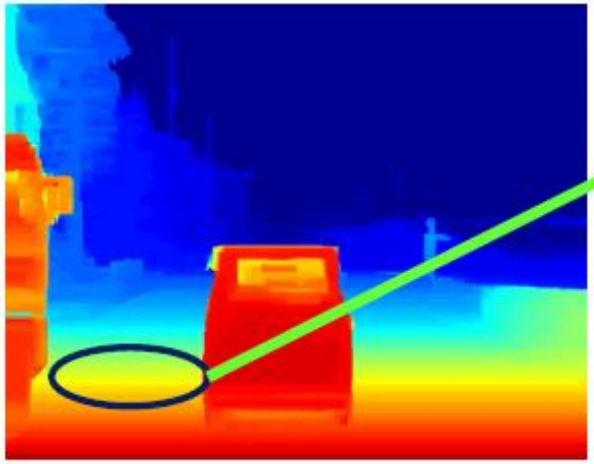
.

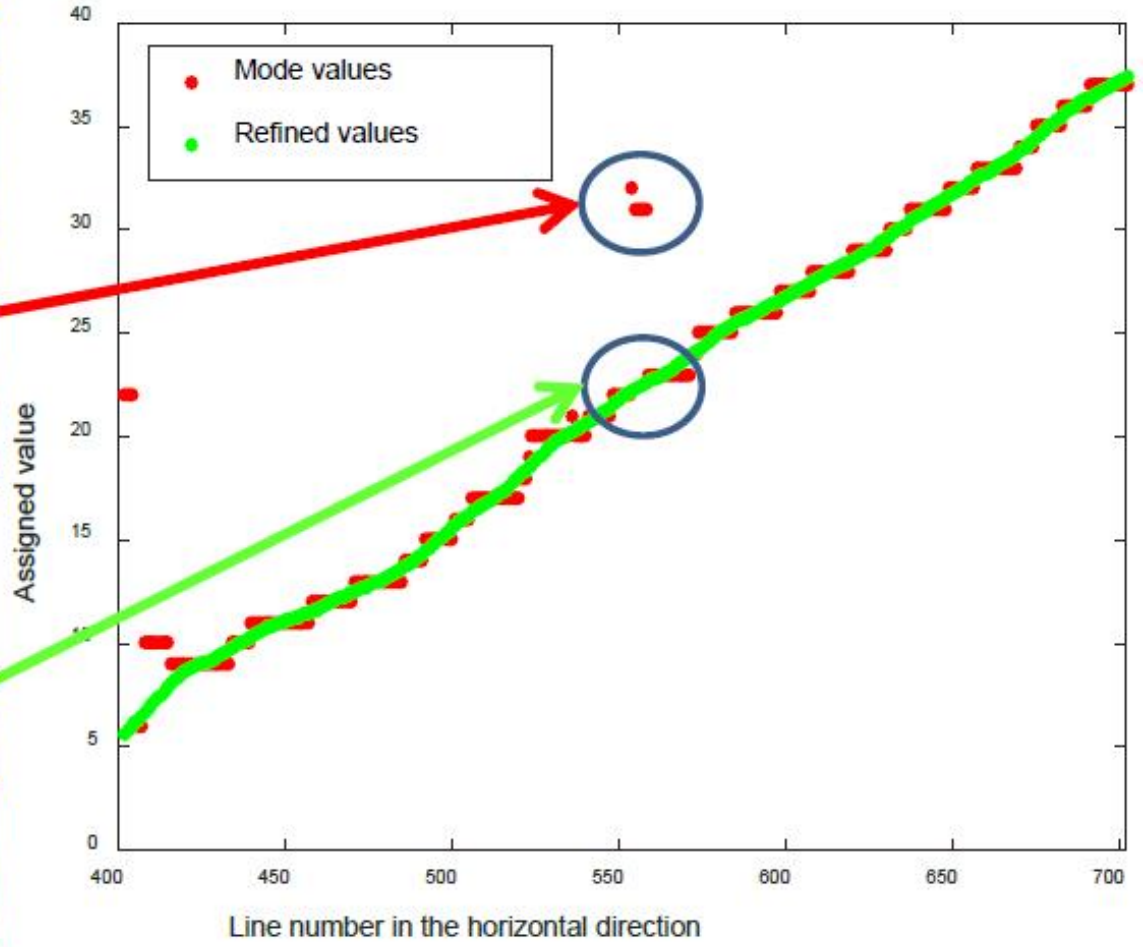
Step 3: Outline of proposed method

- The assignment of the strictly increasing disparity values for each horizontal line is performed using an optimization model.
- As a precursor for the optimization, we first consider the disparity values in each horizontal line and obtain the mode
- Then we detect the abnormal modal indices replaced with the neighboring modal value.

Outline of proposed method







Left: MPV result (up) and road-refined result (down). Right: the original mode values (red points) and refined mode values (green points) in road part.

豊田工業大学
TOYOTA TECHNOLOGICAL INSTITUTE

Step 3: Outline of proposed method

- An optimization model is formulated to obtain the disparity map $\widehat{U}$ as follows:

$$\underset{\widehat{U}}{\operatorname{argmin}} \left\| A \circ \widehat{U} - A \circ U_r \right\|^2 + \lambda \left\| K * \widehat{U} \right\|^2$$

- Where $\circ$ and $*$ denote the component-wise operator and convolution operator, respectively.
- The first term enforces the structural similarity between road-refined U_r and estimated $\widehat{U}$ using the binary matrix A (0 for these abnormal regions).
- The second term ensures the smoothness of $\widehat{U}$ by Laplace operator K, where λ is a regularization parameter for balance.

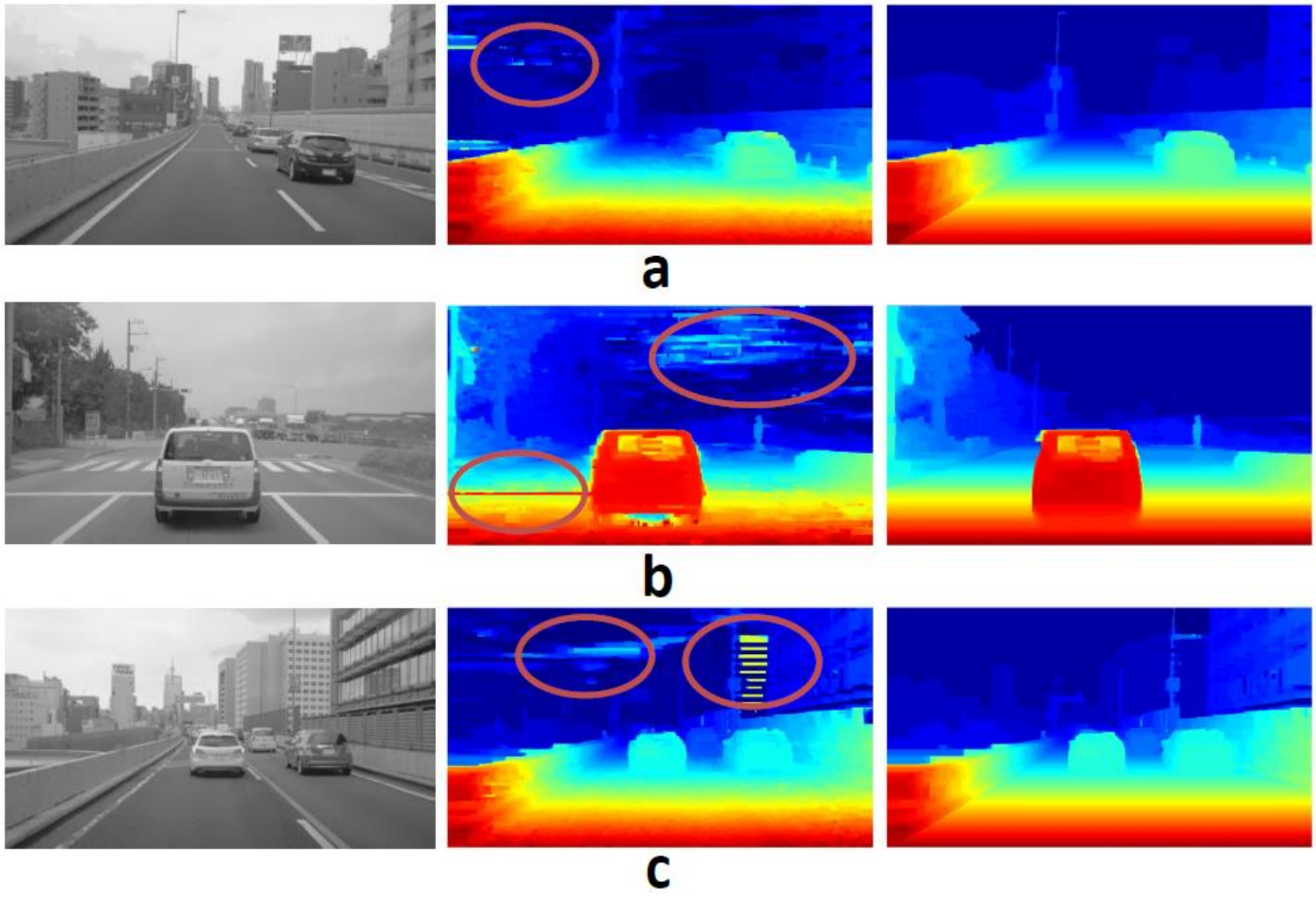
Experimental results

- **Implementation Details**
- The acquired dataset includes 585 frames in several scenes.
- 120 images with the sky and road areas are humanly labeled for training the U-Net.
- During the training process, we use a batch size of 1, and adopt the stochastic gradient descent (SGD) implementation of Keras with a learning rate 0. 01 and decay 0. 0002 as an optimization algorithm.
- We report a processing time of 25ms using Nvidia GPU Geforce 1080.

Experimental results

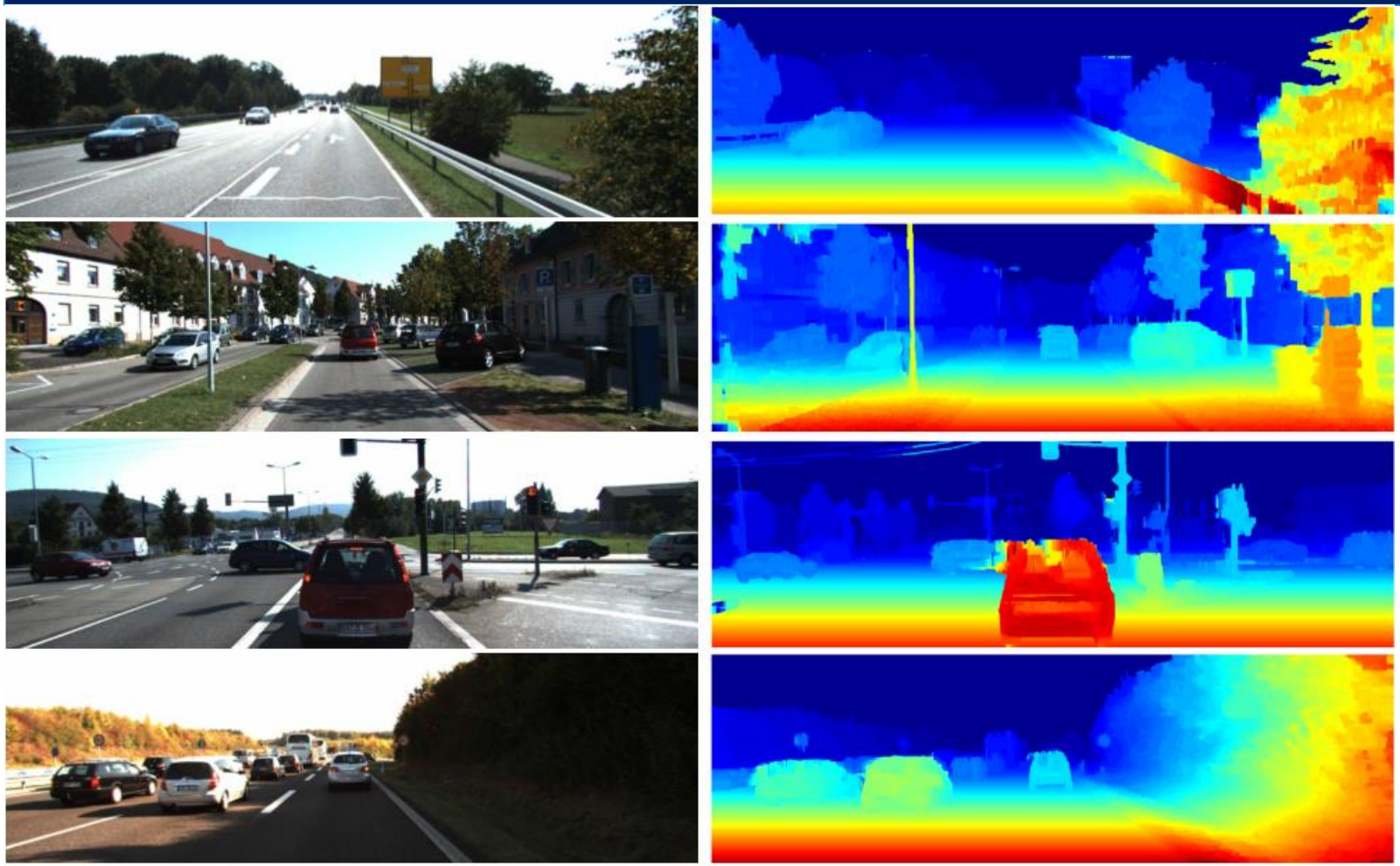
- **Implementation Details**
- Our base MPV algorithm reports a processing time of 40ms in the Geforce TITAN X for 1280 × 960.
- Currently, it is possible to implement the U-Net and stereo vision algorithm in-parallel on our twin GPU Linux machine.
- And the post-processing part is implemented on MATLAB with processing time 2s.

Experimental results



The qualitative results on our dataset. Left: Raw image; Middle: Multi Path Viterbi result; Right: Our results.

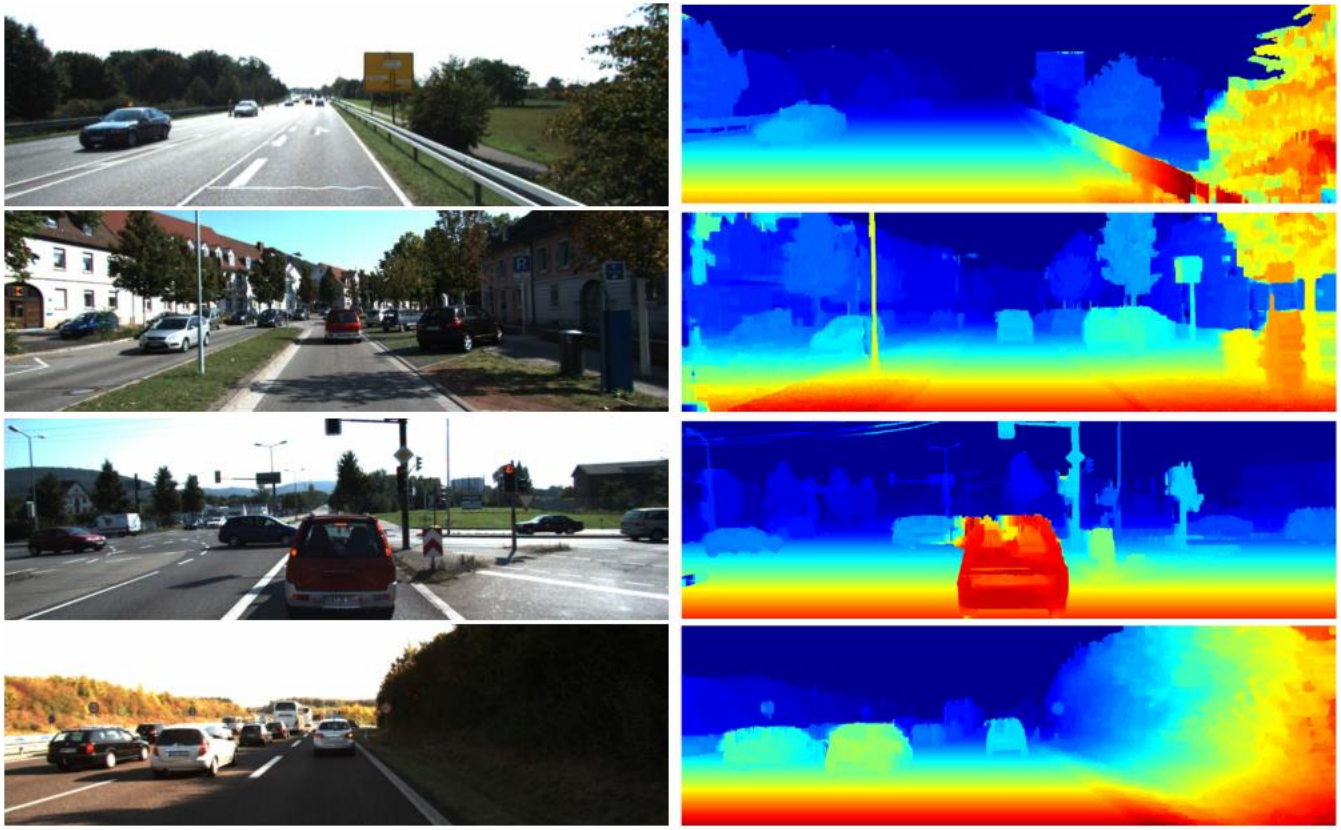
Experimental results



The qualitative results on KITTI dataset. Left: Raw image; Right: Our results.

豊田工業大学
TOYOTA TECHNOLOGICAL INSTITUTE

Experimental results



The qualitative results on KITTI dataset.

Left: Raw image;
Right: Our results.

Table I: Quantitative Analysis of the Proposed Framework.

Algorithm	SGBM	ELAS	MPV	**Proposed**
Average error rate	12.88%	11.99%	7.38%	**7.08%**

豊田工業大学
TOYOTA TECHNOLOGICAL INSTITUTE

Conclusion

- We propose a novel framework to enhance the stereo-based depth generation. Prior information obtained from the U-Net is incorporated within our “base” stereo vision algorithm.
- Multiple refinement models are proposed within the mathematical post-processing framework to refine the depth errors due to the sky, the road and other unknown objects.
- The proposed framework is validated quantitatively and qualitatively on the acquired datasets as well as KITTI dataset.

Thanks

Questions?