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

Weifu Li[1], John Vijay[2], Seiichi Mita[2]
1. College of Science, Huazhong Agricultural University, Wuhan 430062, P. R. China
2. Research Center for Smart Vehicles, Toyota Technological Institute, Nagoya, 468-8511, Japan.

INTRODUCTION

- **Background.** Generation of high density depth values of the driving environment is indispensable for autonomous driving. Stereo vision is one of the practical methods to generate these depth values.
- **Big Challenges**. The accuracy of the stereo vision is limited by texture-less regions, such as sky and road areas, and repeated patterns.
- **Our automated method.** We propose to enhance the stereo generated depth by incorporating prior information of the driving environment.

MAIN CONTRIBUTION

Incorporation of deep learning-based semantic segmentation [1] within the Multi Path Viterbi algorithm [2].

METHODS

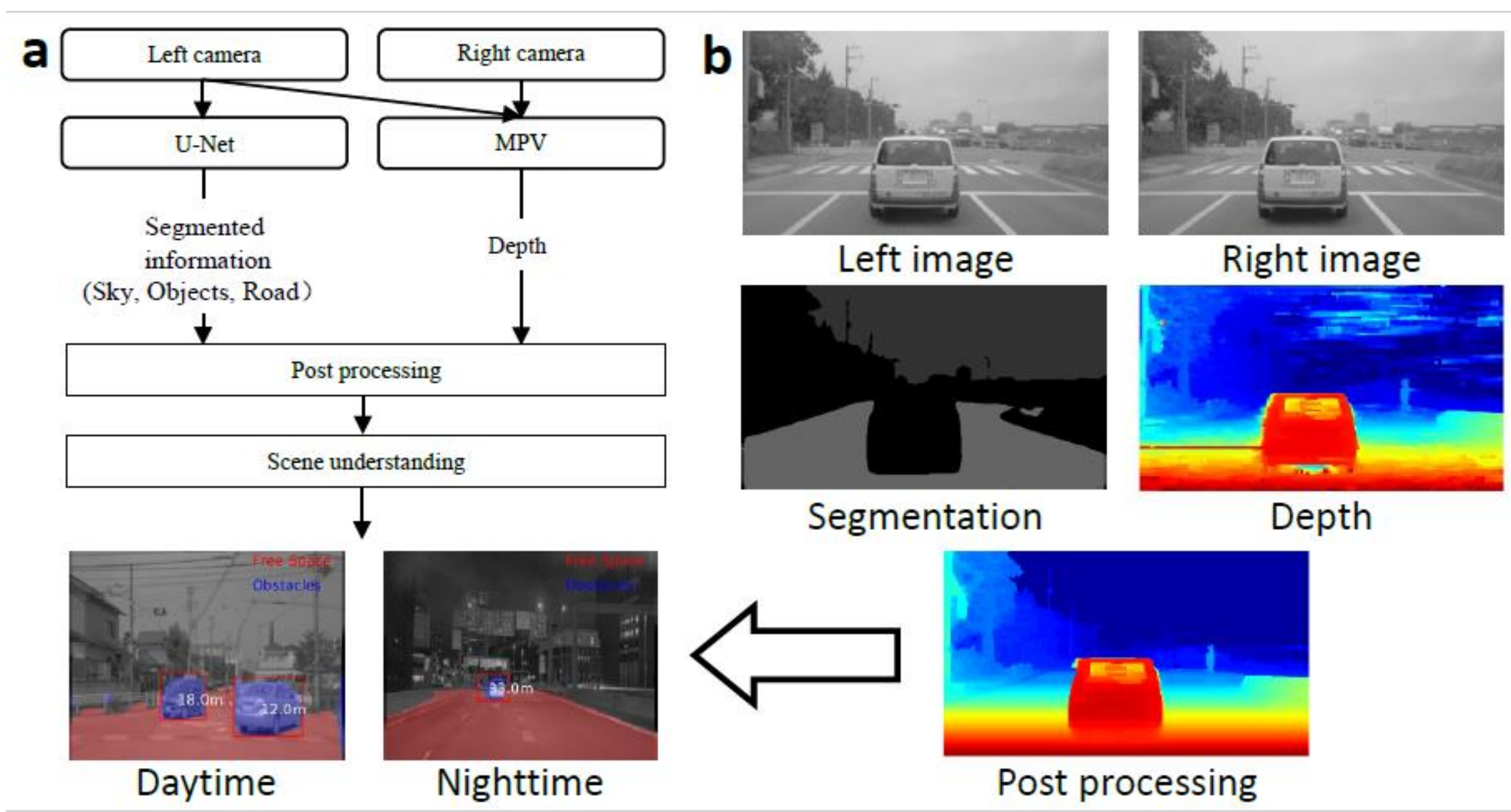

Fig.1 (a) shows the outline of the proposed method. Firstly, depth information for surroundings is generated from the left and right camera images according to the Multi Path Viterbi algorithm. Meanwhile, the same scene obtained from the left camera is segmented into sky, road and other objects. Depth errors of the Multi Path Viterbi algorithm are modified at the post processing steps, as shown in (b). After that, traffic scenes are understood at the scene understanding part shown in (a). Vehicles, pedestrians, obstacles can be detected and distances are estimated using depth values.

Post-processing steps

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.

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.

Other objects. We detect the abnormal regions and replace these regions with the surrounding information. An optimization model is formulated to obtain the disparity map $\hat{U}$ as follows:

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

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 $\hat{U}$ using the binary matrix A (0 for these abnormal regions). The second term ensures the smoothness of $\hat{U}$ by Laplace operator K, where λ is a regularization parameter for balance.

RESULTS

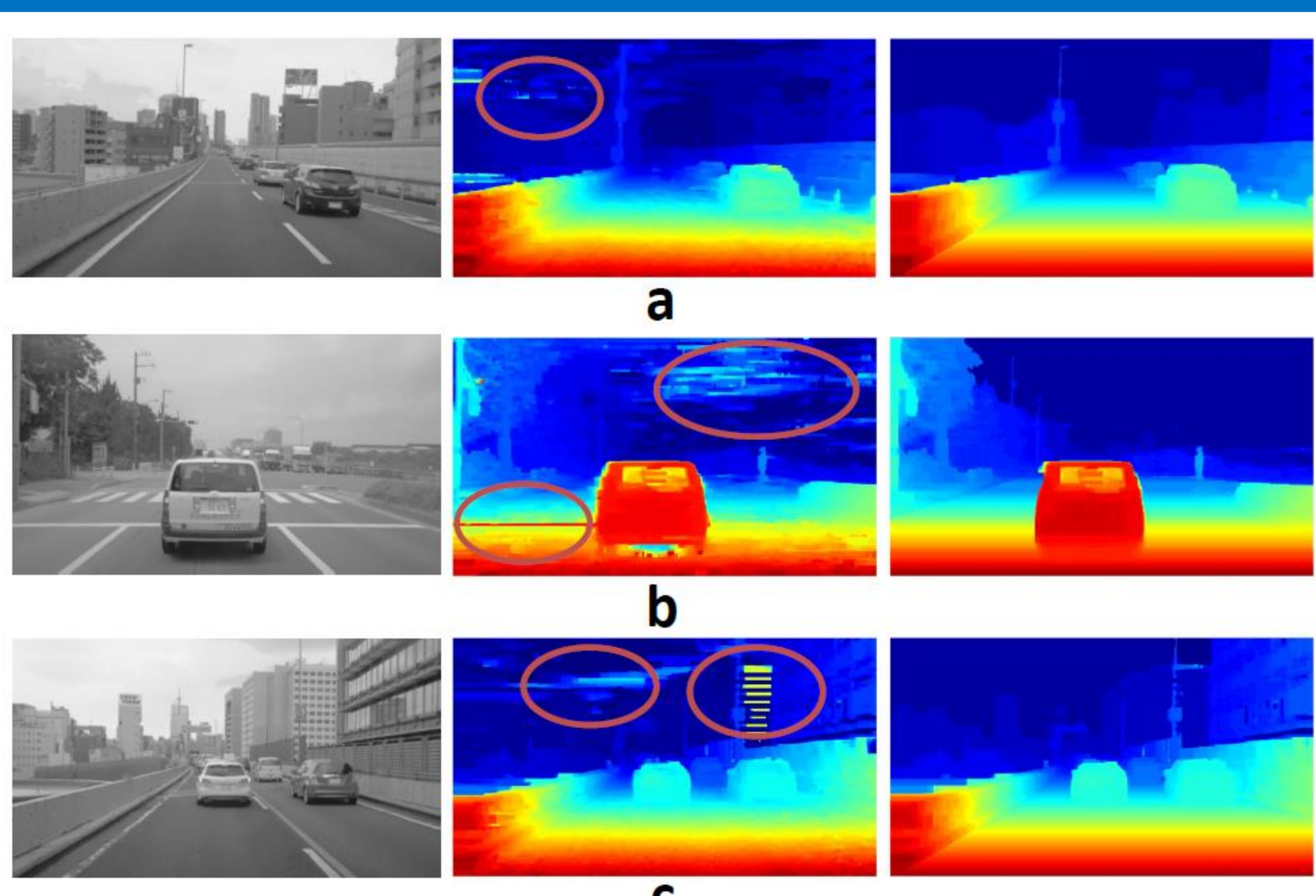

Fig. 2. The qualitative results on our acquired dataset. Left: Raw image; Middle: Multi Path Viterbi result; Right: Our results.

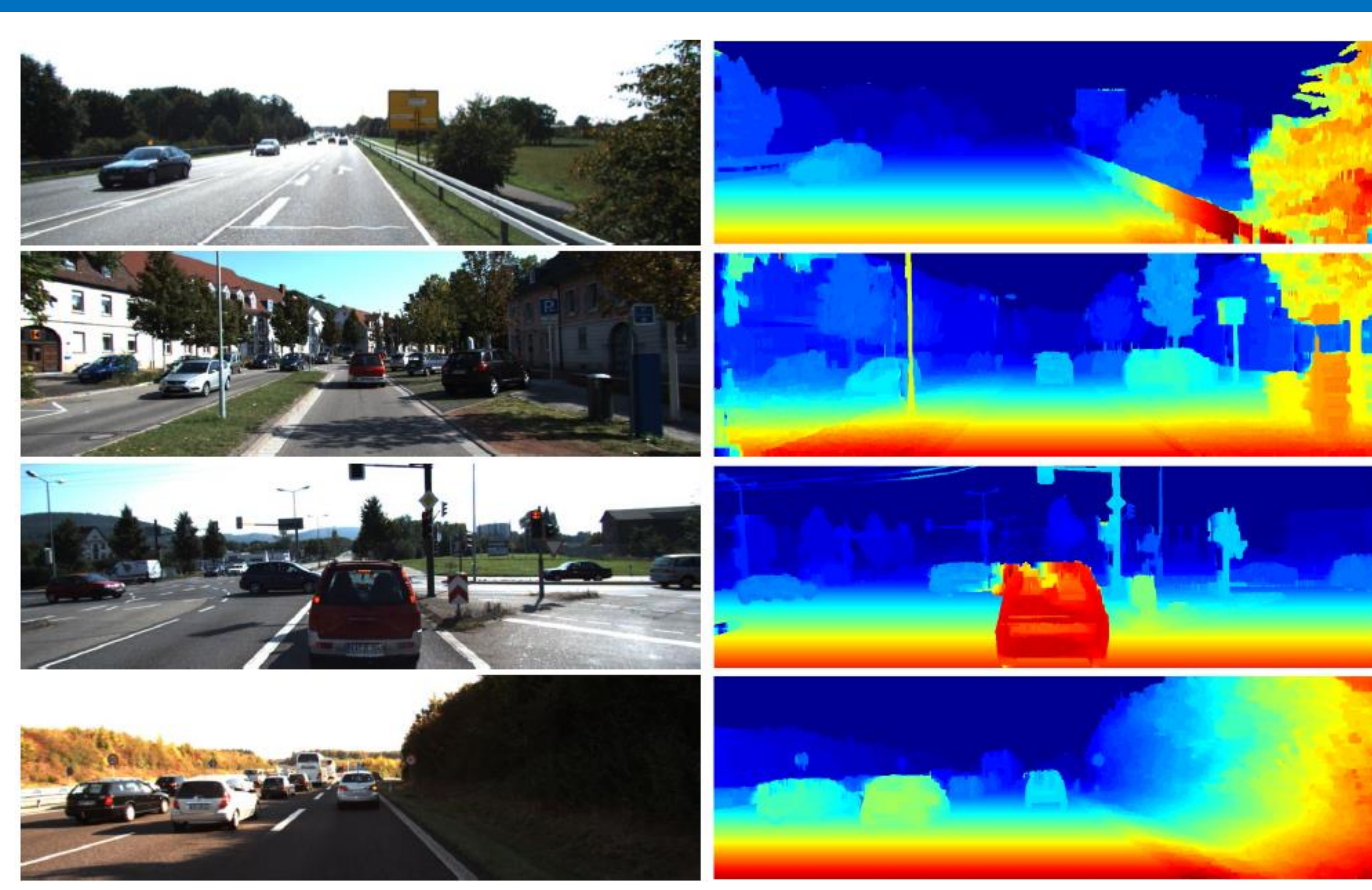

Fig. 3. The qualitative results on KITTI dataset. Left: Raw image; Right: Our results.

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.

REFERENCES

[1] O. Ronneberger, P. Fischer, T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, vol. 9351: 234-241, 2015.

[2] Q. Long et al. “Real-time dense disparity estimation based on multi-path Viterbi forintelligent vehicle applications," in Proceedings of the British MachineVision Conference. BMVA Press, 2014.