DENSE ROAD SURFACE GRIP MAP PREDICTION FROM MULTIMODAL IMAGE DATA

***Jyri Maanpää**[*,1,2] **Julius Pesonen**[*,1,2] Heikki Hyyti[1] Iaroslav Melekhov[2] Juho Kannala[2,3] Petri Manninen[1] Antero Kukko[1] Juha Hyyppä[1]*

[1]Finnish Geospatial Research Institute FGI, National Land Survey of Finland, 02150 Espoo, Finland

[2]Department of Computer Science, Aalto University, 02150 Espoo, Finland

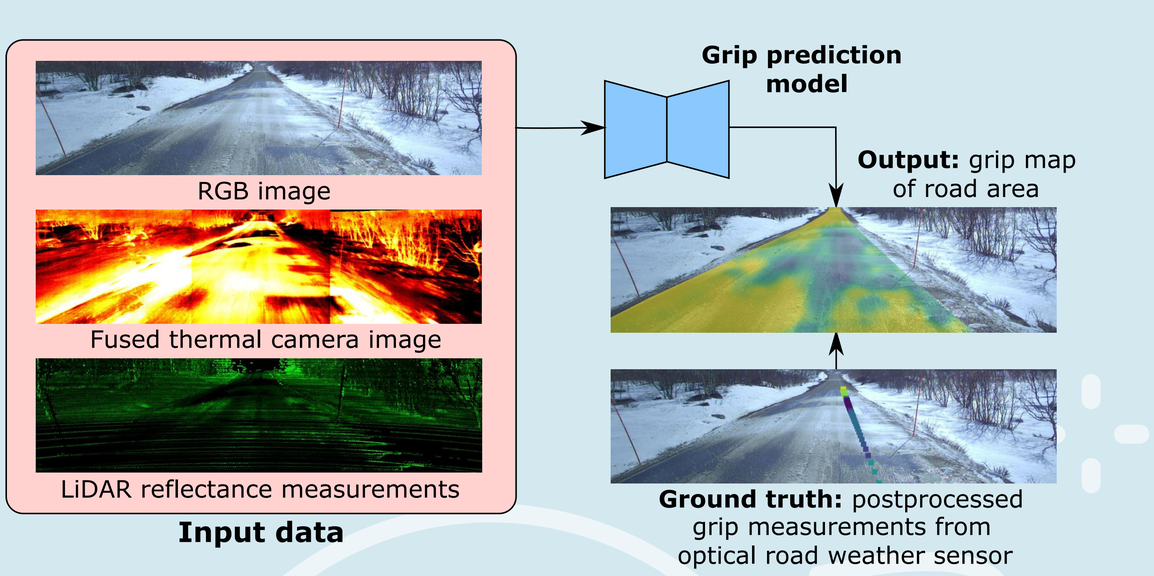
[3]University of Oulu, 90570 Oulu, Finland

*Equal Contribution

ABSTRACT

Slippery road weather conditions are prevalent in many regions and cause a regular risk for traffic. Still, there has been less research on how autonomous vehicles could detect slippery driving conditions on the road to drive safely. In this work, we propose a method to predict a dense grip map from the area in front of the car, based on postprocessed multimodal sensor data. We trained a convolutional neural network to predict pixelwise grip values from fused RGB camera, thermal camera, and LiDAR reflectance images, based on weakly supervised ground truth from an optical road weather sensor.

The experiments show that it is possible to predict dense grip values with good accuracy from the used data modalities as the produced grip map follows both ground truth measurements and local weather conditions, such as snowy areas on the road. The model using only the RGB camera or LiDAR reflectance modality provided good baseline results for grip prediction accuracy while using models fusing the RGB camera, thermal camera, and LiDAR modalities improved the grip predictions significantly.



◄ DATA COLLECTION VEHICLE

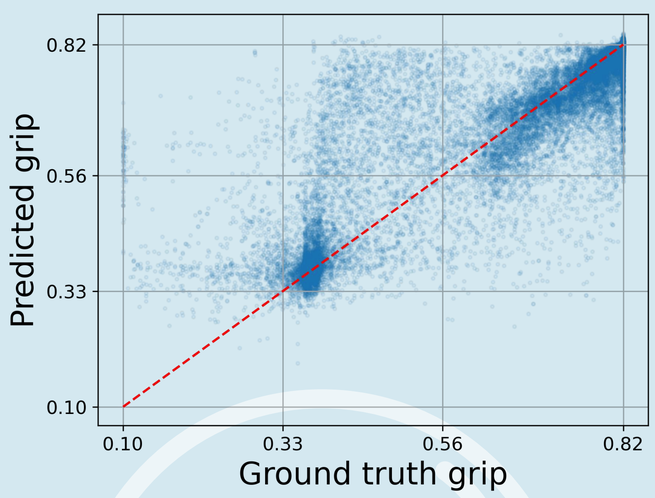
We collected a **37-hour** (1538km, 237 000-sample) dataset with our data collection vehicle. It has data from **Lidar** (A1), **RGB** camera (A2) and three **thermal camera** (A3) sensors in addition to **road weather sensor** (B) data.

The road weather sensor data was matched with other sensor modalities with the GNSS trajectory of the car and the calibrations between different sensors.

▲ OUR MODEL

Our model is a convolutional neural network, operating on **RGB** camera, fused **thermal camera** and/or **lidar reflectance** measurements. The model was trained with grip measurements from an optical road weather sensor, which were matched to the road area during the postprocessing of the collected data. As a result, the model outputs a pixelwise grip map, predicting the grip value for the corresponding input data pixel.

The model is a **Feature Pyramid Network (FPN)** [1] using separate **ResNet18** [2] encoders for different input modalities.



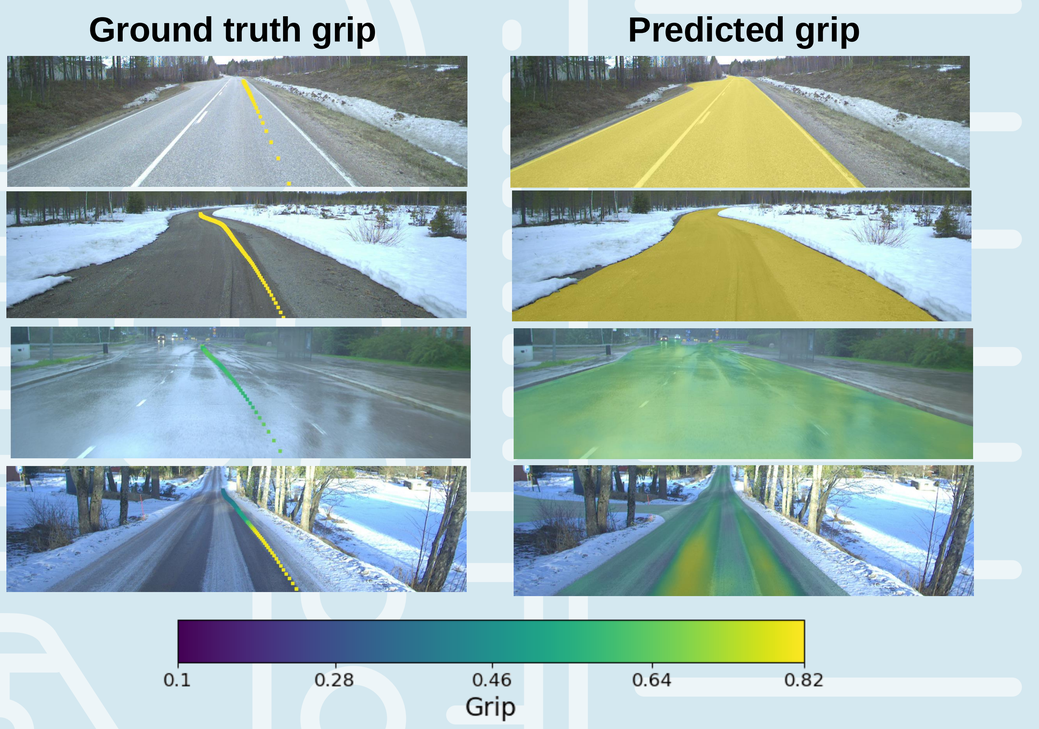
RESULTS

◄ The model learned to predict the road grip values, mostly in dry (grip 0.82) and snowy region (grip 0.4).

The sensor modality comparison ►: shows that the models using RGB camera (**RGB**) and Lidar reflectance measurements (**R**) as input were the most useful in grip prediction as the use of thermal camera (**T**) improved the grip prediction accuracy less.

Model input modalities	Validation set RMSE	Test set RMSE
RGB	0.0657	0.0589
T	0.0794	0.0772
R	0.0677	0.0591
RGB + T	0.0655	0.0605
RGB + R	<u>0.0638</u>	**0.0565**
T + R	0.0664	0.0586
RGB + T + R	**0.0632**	<u>0.0575</u>

▼ The model predicts the grip in several scenarios.



References

[1] Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2117–2125. IEEE (2017)

[2] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778. IEEE (2016)

Acknowledgements

Funded by the European Union (grant no. 101069576). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.