Dear Editorial Board of IEEE Robotics and Automation Letters

We hereby submit the manuscript entitled **"MLPD: Multi-Label Pedestrian Detector in Multispectral Domain"** for consideration by **"IEEE Robotics and Automation Letters"**. We confirm that this work is original and has not been published elsewhere in any format, nor is it currently under consideration for publication elsewhere.

Most multi-modal fusion methods have developed algorithms using the fully-aligned image pair. In multispectral pedestrian detection, KAIST dataset, which is taken with special equipment, such as beam splitter, is the most universally used. However, such complicated sensor configurations that obtain the pixel-level alignment are not easy to apply to real-world applications. In practice, stereo-vision system or EO/IR camera is used in many existing solutions, but this sensor configuration cannot guarantee the pixel-level alignment between multispectral image pair, due to the parallax, sensor resolution, or vehicle vibration. Since this case is completely different from the assumption of the existing fusion methods, the existing methods usually show a large performance drop.

In the manuscript, we tackle the multispectral pedestrian detection, where all input data is not perfectly aligned, and we introduce a generalized multispectral pedestrian detection framework that detects the pedestrian in both paired (fully-aligned) and un-paired (partially-aligned) conditions. The main contents are summarized as follows: 1) We address constraints of previous fusion methods, which make the methods hard to be applied to real world applications and introduce a new perspective of the multispectral pedestrian detection in unpaired conditions; 2) We propose a generalized multispectral pedestrian framework for ideal and practical conditions, which is built upon multi-label learning with a novel augmentation technique; 3) We test the proposed method, considering various unpaired cases, and it achieves the comparable or better result in comparison to the state-of-the-art algorithm.

To demonstrate the validity of the proposed method, we conduct various experiments in multispectral pedestrian datasets, as well as synthetically generated datasets. That is, we conducted extensive experiments regarding various conditions: 1) General paired images; 2) Sensor failure 3) Stereo camera 4) EO/IR camera. The experimental results indicate the superiority of our proposed method in misalignment and unpaired conditions in comparison to the state-of-the-art pedestrian detection algorithm.

We would like to thank you, in advance, for your consideration of the submitted work and are looking forward to receiving your valuable feedback on the manuscript. We hope you will find that our results would be of interest to readers of **IEEE *Robotics and Automation Letters***.

Sincerely,

Jiwon Kim, Hyeongjun Kim, Taejoo Kim, Namil Kim and Yukyung Choi