Sensors and Materials, Vol. 34, No. 7 (2022) 2615–2624
MYU Tokyo

S & M 2994

Multi-sensorial Image Dataset Collected from Mobile Mapping System for Asphalt Pavement Management

JaeKang Lee[1*] and Yong Huh[2]

[1]Korea Institute of Civil Engineering and Building Technology, Koyang-si 10223, Korea
[2]Korea Research Institute for Human Settlements, Sejong-si 30147, Korea

(Received November 13, 2021; accepted January 18, 2022)

Keywords: asphalt pavement management, image registration, mobile mapping system, infrared dataset

In this study, we present a new dataset for managing asphalt pavement surfaces, especially for crack detection. To achieve this goal, we installed a multi-sensor system on the mobile mapping system (MMS) and obtained real-time RGB and IR images, and then the geometric constraint method was applied to find corresponding feature points to spatially register these images. Finally, three environmental data consisting of temperature, humidity, and wind speed are added to the images according to time and location. These data are integrated according to the proposed database model. The proposed system was tested and the databases were constructed for our experiment site, namely, the Capital Region First Ring Expressway in Goyang-si, Gyeonggi Province, South Korea. A total of 800 multi-sensorial images were collected from the expressway. The developed database can be used to train deep learning networks so that it will support detecting road signs or damage on asphalt surfaces.

1. Introduction

Recently, advances in machine learning or deep learning have accelerated the construction of public datasets, such as ImageNet,[1] PASCAL VOC,[2] KITTI,[3] and COCO.[4] However, the public datasets built so far consist of daily images, and there is a lack of data for specific purposes, such as facility diagnosis and road surface condition assessment. Furthermore, because conventional datasets are composed of only visible images consisting of RGB channels, they are not suitable for detecting irregular objects from asphalt surfaces, which are difficult to distinguish on the basis of only color information. Hence, several studies have presented public datasets that provide conventional RGB images and infrared (IR) thermal images simultaneously.[5,6] These datasets consist of images taken by RGB and IR cameras installed on driving vehicles so that they can be used to classify surrounding objects when operating autonomous vehicles. However, there is still a lack of image data that include vertically photographed road surfaces to detect road surface damage more intuitively. The lack of such public datasets has been a major obstacle in developing a technique to detect damage on the asphalt surface using deep learning. In this study, therefore, we install regular cameras and an

*Corresponding author: e-mail: jaekang.lee@kict.re.kr
https://doi.org/10.18494/SAM3731

ISSN 0914-4935
https://myukk.org/

IR camera simultaneously on the bottom part of a mobile mapping system (MMS) to obtain image data. We perform the automatic registration of the image data; then, the sensor data from the MMS's global positioning system (GPS) and an inertial measurement unit (IMU) and the environmental data about temperature, humidity, and wind speed are integrated into the above image data according to location and time.

2. Literature Review

2.1 MMS image dataset

Recently, owing to the advancement of deep learning technology, the need for public datasets for learning and verification of deep learning models has emerged. However, since most public datasets consist of general photos, they are difficult to use for detecting damage on road surfaces using deep learning. To this end, several researchers published road image datasets taken by MMS. Fritsch *et al.* presented a novel open-access dataset and benchmark for a road area collected from MMS, called the KITTI-ROAD dataset.[7] The KITTI-ROAD dataset consists of 600 frames that are extracted from the original KITTI dataset.[3] The dataset also has color stereo images, laser scans, and GPS information. More recently, Maddern *et al.* have introduced a hand-annotated MMS image dataset called Oxford Robotcar Dataset.[8] The Oxford Robotcar Dataset collected from six cameras mounted on the vehicle, along with LIDAR, GPS, and INS. The dataset was used to detect road boundaries.[9] In addition, Brostow *et al.* presented a high-definition video dataset taken from CCTV-style cameras installed in a car, named CamVid.[10] The CamVid dataset contains 32 semantic classes. However, most road image datasets consist of images facing the front of a car, so that they can be used to detect pedestrians and objects around the road for autonomous driving. Therefore, to more precisely detect damage on the road surface, a new image dataset obtained by photographing the road surface vertically is needed.

2.2 IR thermography dataset

Until now, IR thermography datasets have been mainly used to detect and track humans in images.[11,12] Park *et al.* proposed a method that uses a dataset of IR closed-circuit television (IR CCTV) images captured at high angles to train a deep learning network and detect humans.[13] Ligocki *et al.* presented a dataset composed of images recorded by RGB and IR cameras installed on the ceiling of MMS.[14] The dataset consists of 67 video image data captured in different environments and weather conditions, and the total road length in the video recording was 375 km. Krišto and Ivašić-Kos also presented a dataset consisting of 7412 thermal images that captured movements of various people and used it to detect people in the images.[12] The images were captured using long-wavelength IR (LWIR), and like other datasets, they were captured in various environments and weather conditions. As such, most thermal image datasets have been created to train deep learning networks to detect humans. Thus, the optical lenses are facing forward, and consequently, the field of view is inadequate for detecting damage on road surfaces. Golrokh *et al.* proposed a method that combines thermal and visible images captured

perpendicular to road surfaces.[15] The method combines the data taken from stationary sensors, which have a higher quality than the images captured in real time by MMS, but it may not be adequate because the amount of data obtained over time is significantly insufficient. Furthermore, thermal images may vary greatly depending on the surrounding environmental factors such as temperature and humidity. In this study, therefore, we obtain real-time images while driving through a sensing system installed at the bottom part of the MMS and divide them into time-series image data to enhance the quality of data. Table 1 shows the comparison of the existing MMS datasets.

2.3 Registration of RGB and IR images

Various methods to geo-register RGB and IR images have been developed so far. Various studies have used, as the most simple and intuitive method, the registration of two images by calculating the transformation parameters between images when the internal/external parameters of cameras and the phase difference between sensors are known.[16] An image registration method of the images using an image feature matching method has also been used in various studies.[17–19] Recently, as the development of machine learning methods has accelerated, registration methods of images obtained from multiple sensors have been proposed.[20–22] These methods train the model by inputting RGB and IR images simultaneously to extract semantic information, on the basis of which image registration is performed. However, although the image registration methods described above work well when the scale or field of view is very similar between the images, it is difficult to perform the image registration when they are geometrically deformed or distorted. Particularly, in the case of cameras mounted on the lower part of the MMS, images are often distorted and out of focus, and consequently, there are many difficulties in applying the above methods as they are.

Therefore, in this study, we use the geometric constraint method that randomly selects three matching point pairs in the two images and measures the ratio of the corresponding line segment lengths according to the coordinates of the three feature points.[17] On the basis of the ratio, the three most similar feature point matching pairs are obtained to be used for affine geometric deformation to perform the image registration.

Table 1
Comparison of MMS datasets.

		KITTI-ROAD	Oxford Robotcar Dataset	CamVid	Our dataset
# of images		600	62605	700	800
Image resolution (IR camera)		800 × 400	1280 × 960	960 × 720	1624 × 1236 (640 × 512)
Sensing data	RGB	○	○	○	○
	LiDAR	○	○	×	×
	IR	×	×	×	○
	GPS	○	○	×	○
	Weather condition	×	×	×	○

3. Dataset

3.1 Data acquisition equipment

In this study, the data were obtained through the scalable sensor platform attached to the lower part of the MMS equipment. The sensing system consists of RGB and IR sensors, GPS, and IMU, and the real-time data transmission is facilitated through network communication equipment. Furthermore, environmental factors such as temperature, humidity, and wind speed are obtained through the network communication equipment, and the time is synchronized between data through a GPS clock. Figure 1 and Table 2 show the sensing system used in this study.

3.2 Data acquisition section

The test bed considered for this study is a 9 km section on the Capital Region First Ring Expressway in Goyang, Gyeonggi Province as shown in Fig. 2. The expressway is a circular

Fig. 1. (Color online) Multi-sensor system installed on MMS.

Table 2
Sensors used in this study.

	Type	Resolution	Output data
2 × RGB camera	Lynx camera	1624 × 1236	Three-channel (RGB) image
IR camera	FLIR Tau 2	640 × 512	One-channel image
GNSS/INS	Pos LV420	*XY*: 0.02m *Z*: 0.05m	Three-dimensional coordinates

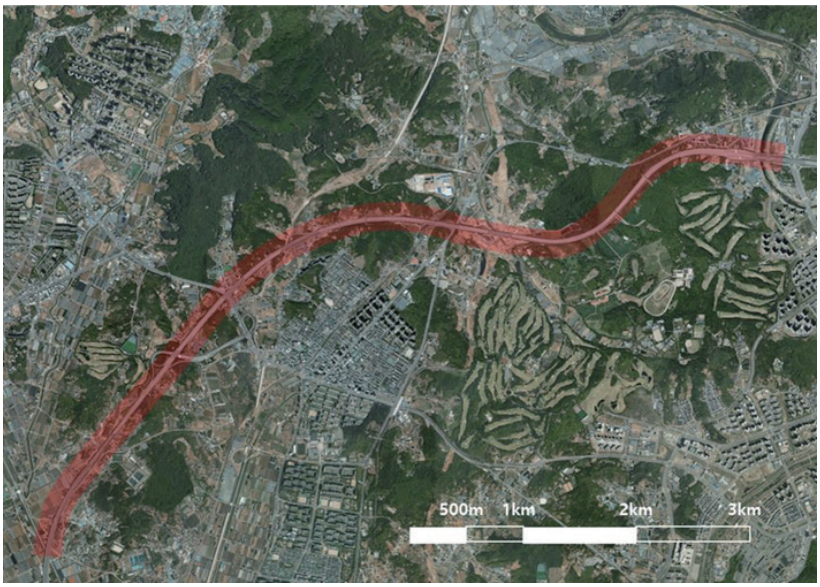


Fig. 2. (Color online) Experimental area in this study.

beltway around Seoul that connects satellite cities around Seoul, including Goyang. Since the experimental section has been built for over 20 years, precise management of the pavement is required.

3.3 Data composition

As described above, the dataset consists of RGB and IR video recording data, sensor data, and environmental data about temperature, humidity, and wind speed. The video data are divided into frame images of specific time according to the GPS time. Therefore, a pair of data consists of an RGB image, an IR image, environmental data, and location information data. In this study, the video dataset consists of GPS and IMU sensing data and RGB and IR video data; the image metafile dataset consists of the images constructed by dividing the video into one frame for a second, which are synchronized with the environmental data. Here, the values of temperature, humidity, and wind speed, which are environmental data, are stored in units of °F, %, and m/s, respectively. Figure 3 shows sample data of the image dataset.

3.4 Construction of registered metafile and dataset

In this study, we constructed a metafile by registering RGB and IR images and environmental data to provide more features to machine learning. The thermal and visible images output in the sensing system have different fields of view because they were captured at different positions, and the image resolution and internal expression parameters of cameras are also different. Furthermore, because the images were captured from the lower part of the MMS while driving, many images were distorted geometrically. In this study, therefore, we used a multi-modal image registration method to register the images outputted from different sensors. First, in the output

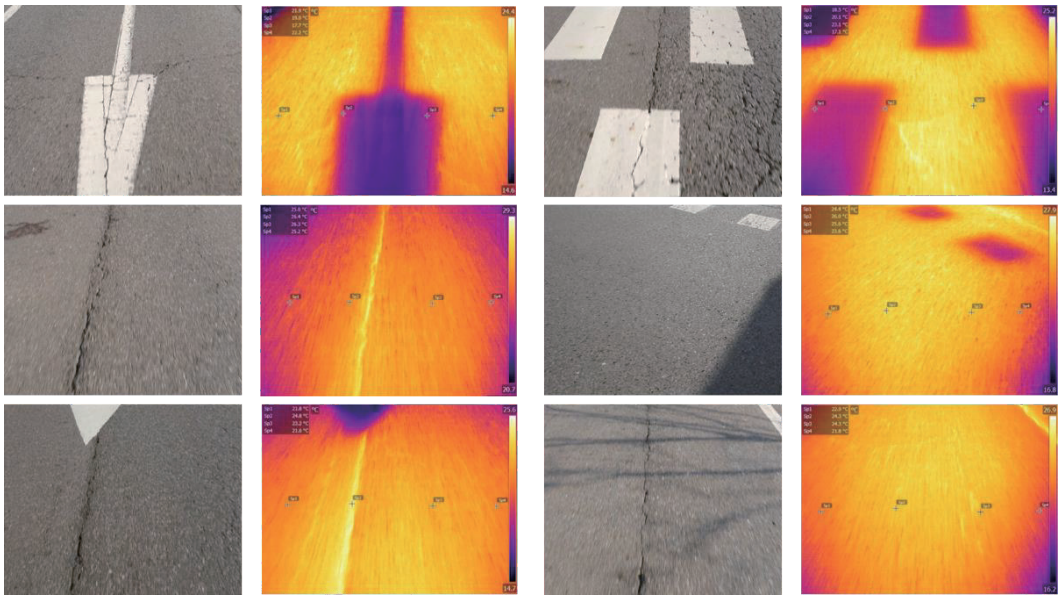


Fig. 3. (Color online) RGB image and IR image data output.

image pair, the RGB image is designated as a reference map, and the image is registered through the affine transformation of the IR image. The following equation shows the affine transformation:

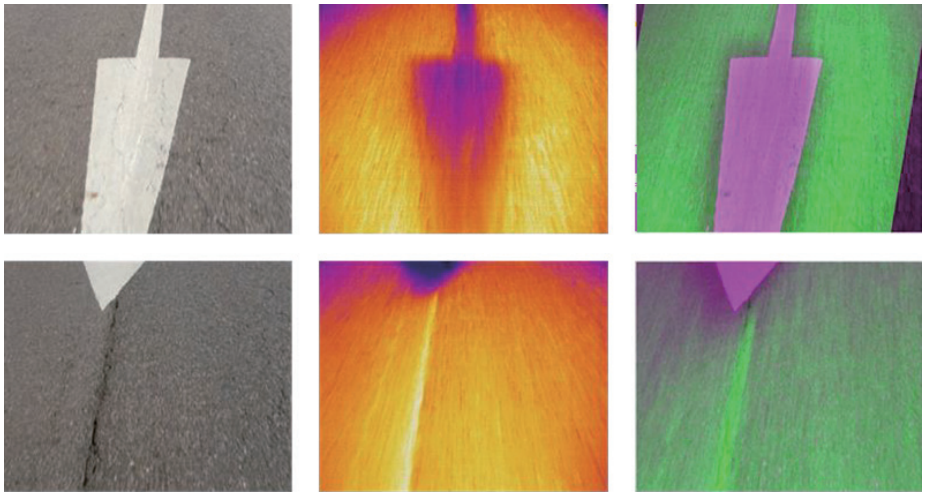
$$P_{fixed} = T \cdot P_{moving}, \tag{1}$$

where P_{fixed} denotes the pixels of the RGB image, and P_{moving} denotes the pixels of the IR image. T denotes the affine transformation matrix, which is calculated as follows:

$$T = R(\theta) \cdot S_h(k) \cdot S(s_x, s_y) \cdot T(t_x, t_y), \tag{2}$$

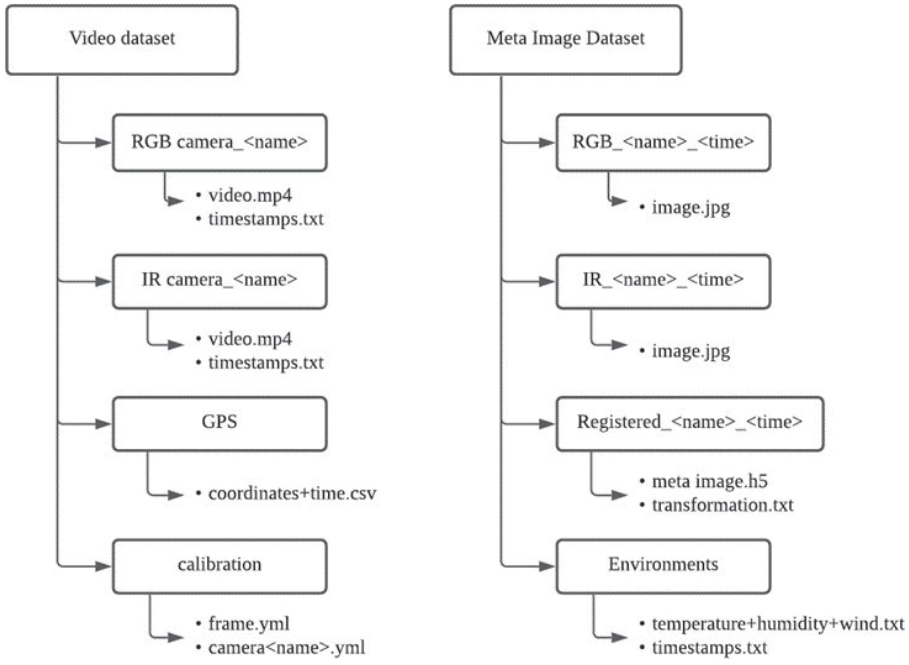
where $R(\theta)$ denotes the rotation matrix at an angle θ, $S_h(k)$ denotes the shearing transformation, $S(s_x, s_y)$ denotes the scale transformation for the x- and y-axes, and $T(t_x, t_y)$ denotes the translation. Figure 4 shows the registered image metafile.

The registered metafile is saved in the final form: three environmental data consisting of temperature, humidity, and wind speed are added to four channels, in which the IR intensity value is added to the RGB channels, forming seven channels in total. As described earlier, the dataset developed in this study is provided in two main forms (videos + sensing data, meta-image file). First, the video dataset is classified into recording entity units, each stored in a separate folder. Figure 5(a) shows the structure of the video dataset. In the case of the meta-image dataset, the RGB and IR images are captured at the same time as registered image and environmental data from a data set, respectively. Figure 5(b) shows the structure of the meta-image dataset. Each individual dataset is compressed and provided in a .zip format. As described earlier, the registered meta-image file is saved as an h5 file that has seven channels.



(a) (b) (c)

Fig. 4. (Color online) Meta-image dataset samples. (a) RGB images, (b) IR images, (c) registered images.



(a) (b)

Fig. 5. Data structure of (a) video dataset and (b) meta-image dataset.

In this study, we built an online database containing the datasets as shown in Fig. 6. The image data stored in the online database are visualized by time and section, and in addition, the environmental data and image processing results are also shown through visualization. The visualized image processing results show the RGB histogram for each section and the maximum/minimum temperature in the IR image.

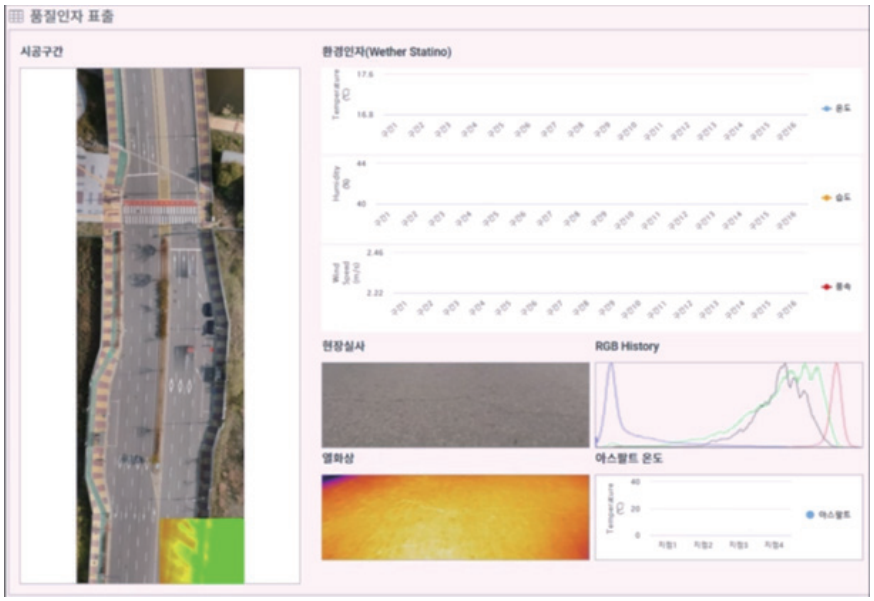


Fig. 6. (Color online) Online database.

4. Conclusions

In this study, we presented a new dataset for detecting damage on asphalt pavement surfaces. To achieve this goal, we installed a multi-sensor system on the MMS to record 9 km of the road surfaces while driving, and we synchronized the time between the multiple sensors using the GPS clock. The dataset provided in this study is mainly classified into a video dataset and a meta-image dataset. First, the video dataset provides the RGB camera and IR camera recordings and corresponding GPS data. The meta-image dataset, on the other hand, provides the basic RGB images and IR images, and the multi-channel registered images for the deep learning network training. The RGB and IR images were registered after performing affine transformation using a multi-modal non-rigid method. The dataset may be used to train deep learning networks, such as YOLO detectors, and we expect that it will be used mainly to detect road signs or damage on asphalt surfaces. The limitation of this study is that it provides image data for asphalt surfaces only. In a future study, therefore, we plan to obtain image data for concrete pavement surfaces and provide a registered dataset. Furthermore, we will provide a label file according to the COCO annotation style to increase the compatibility of the dataset with various deep learning networks.

Acknowledgments

This work was supported by the Korea Institute of Civil Engineering and Building Technology Research Project (Research on Smart Construction Technology for Leading the future construction industry and Creating new market) (2022).

References

1 O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. Berg, and L. Fei-Fei: Int. J. Comp. Vis. **115** (2015) 211. https://doi.org/10.1007/s11263-015-0816-y
2 M. Everingham, S. Ali Eslami, L. van Gool, C. Williams, J. Winn, and A. Zisserman: Int. J. Comp. Vis. **115** (2015) 98. https://doi.org/10.1007/s11263-014-0733-5
3 A. Geiger, P. Lenz, and R. Urtasun: Proc. 2012 IEEE Conf. Computer Vision and Pattern Recognition (IEEE, New York, 2012) 3354–3361. https://doi.org/10.1109/CVPR.2012.6248074
4 T. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. Zitnick: Microsoft COCO: Common Objects in Context, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. (Springer Cham, New York, 2012). https://doi.org/10.1007/978-3-319-10602-1_48
5 Y. Choi, N. Kim, S. Hwang, K. Park, J. Yoon, K. An, and I. Kweon: IEEE Trans. Int. Trans. Syst. **19** (2018) 934. https://doi.org/10.1109/TITS.2018.2791533
6 A. Ligocki, A. Jelínek, and L. Žalud: Proc. 2020 IEEE ICRA (IEEE, New York, 2020) 3284–3290. https://doi.org/10.1109/ICRA40945.2020.9197277
7 J. Fritsch, T. Kuehnl, and A. Geiger: In 16th Int. IEEE Conf. Intelligent Transportation Systems (ITSC 2013) 1693–1700.
8 W. Maddern, G. Pascoe, C. Linegar, and P. Newman: Int. J Robot Res. **36** (2017) 3. https://doi.org/10.1177/0278364916679498
9 T. Suleymanov, M. Gadd, D. De Martini, and P. Newman: arXiv (2021) arXiv:2106.08983.
10 G. Brostow, J. Fauqueur, and R. Cipolla: Pattern Recogn. Lett. **30** (2009) 88. https://doi.org/10.1016/j.patrec.2008.04.005
11 C. Li, X. Liang, Y. Lu, N. Zhao, and J. Tang: Pattern Recognit. **96** (2019) 106977. https://doi.org/10.1016/j.patcog.2019.106977
12 M. Krišto and M. Ivašić-Kos: Proc. 42nd Int. Conf. MIPRO (IEEE, New Work, 2019) 1126–1131. https://doi.org/10.23919/MIPRO.2019.8757208.
13 J. Park, J. Chen, Y. Cho, D. Kang, and B. Son: Sensors **20** (2020) 34. https://doi.org/10.3390/s20010034
14 K. Ivašić, M. Krišto, and M. Pobar: UNIRI-Thermal image dataset-UNIRI-TID (2020).
15 A. Golrokh, X. Gu, and Y. Lu: J. Perform. Constr. Facil. **35** (2021). https://doi.org/10.1061/(ASCE)CF.1943-5509.0001557
16 P. Kim, J. Chen, and Y. Cho: Int. J. Intell. Robot. Appl. **1** (2017) 243. https://doi.org/10.1007/s41315-017-0023-9
17 J. Wang, K. Cheng, C. Kel, M. Liu, L. Cai, and H. Shi: J. Phys. Conf. Ser. **1419** (2019) 012008.
18 J. Park, P. Kim, Y. Cho, and Y. Fang: 18th Int. Conf. Construction Applications of Virtual Reality (2018).
19 J. Dou, Q. Qin, Z. Tu, X. Peng, and Y. Li: Proc. 2016 Chinese Control and Decision Conf. (2016) 5420–5424. https://doi.org/10.1109/CCDC.2016.7531966
20 Y. Fu and X. Wu: Proc. 25th Int. Conf. Pattern Recognition (2021) 10675–10680. https://doi.org/10.1109/ICPR48806.2021.9412293
21 Z. Zhao, S. Xu, C. Zhang, J. Liu, J. Zhang, and P. Li: Proc. 29th Int. Joint Conf. Artificial Intelligence (2020) 970–976. https://doi.org/10.24963/ijcai.2020/135
22 A. Dalca, E. Yu P. Golland, B. Fischl, M. Sabuncu, and J. Iglesias: Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention (2019). https://doi.org/10.1007/978-3-030-32248-9_40

About the Authors

Jae Kang Lee received his Ph.D degree from the University of Nottingham, UK, in 2015. Since 2019, he has been working as a senior researcher at the Korea Institute of Civil Engineering and Building Technology (KICT). His research interests range from safety policy in civil engineering to software programming in relation to GNSS positioning, UAV photogrammetry, and application of the big data platform to civil engineering.

Huh Yong is an associate research fellow in the Geospatially Enabled Society Research Division of the Korea Research Institute for Human Settlements. He received his B.S. and Ph.D. degrees from Seoul National University, Rep. of Korea, majoring in civil engineering and geospatial information. His research interests range from government policy to software programming in relation to image processing, pattern recognition, and data mining.