HSI-Drive: A Dataset for the Research of Hyperspectral Image Processing Applied to Autonomous Driving Systems

K. Basterretxea[1], V. Martínez[2], J. Echanobe[2], J. Gutiérrez-Zaballa[2], I. del Campo[2]

***Abstract*— We present a structured dataset for the research and development of automated driving systems (ADS) supported by hyperspectral imaging (HSI). The dataset contains per-pixel manually annotated images selected from videos recorded in real driving conditions that have been organized according to four environment parameters: season, daytime, road type, and weather conditions. The aim is to provide high data diversity and facilitate the automatic generation of data subsets for the evaluation of machine learning (ML) techniques applied to the research of ADS in different driving scenarios and environmental conditions. The video sequences have been captured with a small-size 25-band VNIR (VisibleNearInfraRed) snapshot hyperspectral camera mounted on a driving automobile. The current selection of classes for image annotation is aimed to provide reliable data for the spectral analysis of the items in the scenes; it is thus based on material surface reflectance patterns (spectral signatures). It is foreseen that future versions of the dataset will also incorporate alternative dense semantic labeling of the annotated images. The first version of the dataset, named HSI-Drive v1.0, is publicly available for download[3].**

I. INTRODUCTION

In the field of autonomous driving and advanced driving assistance systems (ADS and ADAS), perceiving the surrounding environment and extracting meaningful information is a key task. The majority of current image-based object analysis systems for ADS rely on RGB imaging for object detection and tracking [1]. Although state of the art object tracking systems have evolved considerably, there still remain certain robustness issues related to changing weather and illumination conditions, as well as to challenging driving scenarios with rapid changes in target appearance and multiple occlusions between different objects [1], [2]. In ADS, robustness of image analysis is critical, since unreliable systems may produce risky driving situations and possibly even fatal accidents.

The use of hyperspectral sensors can help to improve scene understanding and object tracking system robustness, since they provide richer information about materials than conventional cameras. This extra information can be used to better separate objects and backgrounds, improve object tracking robustness and enhance the performance of image segmentation algorithms. HSI is already being widely applied in some areas such as remote sensing in geoscience and, more recently, in precision agriculture, medical imaging and others [3]. However, the use of HSI in application areas that require autonomy of operation and easy integration in moving platforms has been limited by traditional scanning sensor setup requirements and also by restrictions in the processing power of the accompanying computing hardware. To-day, with the advent of new small-size snapshot cameras that can provide hyperspectral images at the video rates [4], it is possible to setup a HSI system in almost any moving platform, and particularly in ground vehicles. Consequently, how to efficiently process hyperspectral information onboard a moving platform is currently an active research topic.

This new HSI technology has begun to attract the attention of some ADS researchers due to its application potential and downward price trends. However, it is still necessary to carry out a concerted research effort to transfer and adapt this technology to the development of more efficient ADS. One of the key resources necessary for the development of HSI-based algorithms for ADS is undoubtedly the availability of specific datasets containing video sequences and annotated images acquired with modern snapshot HSI cameras. Due to the present shortage of such a resource, the Digital Electronics Design Group (GDED) of the University of the Basque Country, currently involved in the research of the edge processing of HSI with low latency requirements, has started to develop of an extensive, structured database of hyperspectral images acquired with a VNIR small-size snapshot camera. In this paper, we describe the state of development of this project and the details of the first version of this dataset, the HSI-Drive v1.0, which is already publicly available for researchers in the field [3].

II. RELATED WORK

A. Multispectral and hyperspectral imaging in ADS

In order to increase the robustness of ADS systems in changing environmental conditions, some researchers have proposed the additional use of images taken outside the visible spectrum. Since light with longer wavelengths is scattered to a lesser extent, the information received at infrared frequencies has different characteristics from that of the visible ones, and also provides greater range under various conditions. In the field of computer vision for ADS several studies have already explored the use of both near infrared (NIR) [5], [6] and far-infrared (FIR) images [7]-[10]. The use of the FIR spectrum has been mainly applied to the detection of pedestrians, since it fundamentally obtains information on the temperature of objects. Its detractors,

[1]K. Basterretxea is with the Department of Electronics Technology, University of the Basque Country, 48013 Bilbao, Spain koldo.basterretxea@ehu.eus

[2]Researchers are with the Department of Electricity and Electronics, University of the Basque Country, 48940 Leioa, Spain

[3]https://ipaccess.ehu.eus/HSI-Drive

*This work was supported by the Basque Government under grant PIBA-2018-1-0054 and partially supported by the University of the Basque Country under grant GIU 18/122

however, allege that these systems turn out to be inefficient in general, since in summer the temperature differences with the environment are small, and in winter the use of thick layers of clothing does not allow such differences to be detected [5].

A more recent study analyzes the fusion of images in RGB, NIR, MIR (medium frequency infrared) and FIR for the detection and classification of objects by mounting a system of four cameras on a cart to simulate a driving situation [11]. The study shows that the combined information from the four cameras allows better differentiation of the different types of objects for which the system (YOLO) is trained. In [12] the authors describe a lighter CNN (Convolutional Neural Network) architecture for the real-time semantic image segmentation of street scenes acquired with a combination of RGB and thermal images. They show that the segmentation accuracy is increased by adding thermal information and that the algorithm can produce outputs with low latency for ADS when processed on high-end GPUs. All the above-mentioned works generally propose making use of multiple cameras with different sensitivity spectra to add "visual" information complementary to that of the visible spectrum. The applied detection and segmentation algorithms, however, remain identical or very similar to the techniques applied to visible imaging, which makes such complex processing pipelines even heavier to compute.

More recently, some researchers have started to explore the applicability of HSI cameras to the field of autonomous driving. The underlying idea is that the incorporation of richer spectral information can provide a distinct spectral fingerprint for each entity in the image. In addition to increasing the robustness of the detection systems, this approach could potentially allow for the application of lighter algorithms than those used for visible images (more information = less processing). One of the earliest studies in this regard, although not specifically directed at ADS, draws interesting conclusions regarding the ability of images taken with hyperspectral cameras to detect people in urban environments [13]. In this work, a hyperspectral camera (225 bands in a 400 to 2500 nm range) is used to determine the capability of the system by analyzing the spectral separability. Basically the work concludes that the system is indeed capable of significantly improving the discrimination capacity in comparison with the simple use of the RGB bands, but more interestingly, it also concludes that the use of the VNIR spectrum (up to 1000 nm) offers similar results to using the full spectrum (up to 2500nm). This is very relevant to the present proposal since most current low-cost HSI snapshot cameras do not offer spectrometric information beyond 1000 nm. An active research group in the investigation of the use of HSI in ADS is the Active Vision Group (AGAS) of the University of Koblenz-Landau. They have published some papers on this topic reporting interesting results on image segmentation and terrain classification applied to images combining VIS and NIR spectrum information from low-cost snapshot cameras [14]-[16]. Apart from a few very recent foundational works [17], [20], little has been published on this specific topic.

B. HSI datasets for ADS research

At the time of the launch of this project there was only one HSI dataset specifically designed for ADS development: the Hyko dataset created by the above mentioned AGAS group. In the meantime, as far as we are aware, two more datasets have been reported.

Hyko1: In 2017 Winkens et al. [18] presented Hyko, an annotated HSI dataset collected with both XIMEA VIS (470 to 630 nm) and NIR (600 to 975 nm) snapshot cameras mounted on a moving car. Hyko1 contains 233, 25-band NIR images and 280, 15-band VIS images for terrain classification. Annotation masks were generated by per-pixel labeling with five classes: "undefined", "drivable", "rough", "obstacle" and "sky".

Hyko2: Contains 78, 25-band NIR images and corresponding annotated masks with semantic classes for urban scenes (11 classes) and masks with spectral reflectance labels (9 classes). It also contains 163 15-band VIS images and corresponding annotated masks with dense drivability labels (5 classes) and masks with spectral reflectance labels.

Hyperspectral City v1.0: In 2019 You et al. [19] presented Hyperspectral City, a dataset and benchmark for urban autonomous driving scenes. Images were captured with a LightGene camera sensor, which provides 129 spectral channels in the 450 to 950 nm range. This high spatial and spectral resolution sensor produces cubes larger than 1GB. All images were taken over three days in June in varied urban settings and lighting conditions. Car driving speed was in the range of 20-50 Km/h with the camera working at 1 fps. The dataset contains a 367-image training dataset and a 58-image testing dataset, but only 300 images have been labeled. Annotation is focused on semantic segmentation with 10 classes, using coarse labeling for the training set and fine labeling for the testing set.

HSI Road: In 2020 Lu et al. [20] presented HSI road, a HIS dataset for road segmentation. It comprises images taken both in urban and rural scenes. It contains 3,799 scenes with RGB and NIR bands as well as their respective annotation masks. NIR images were captured with a 15-band Ximea camera (ranging from 680 to 960 nm). Image annotation has been created manually by polygon labeling tools. Only two classes are available: "background" and "road".

III. THE HSI-DRIVE DATASET

A. System Setup

The recording system setup for this project was extremely simple, consisting of just one Photonfocus MV1-D2048x1088-HS02-96-G2 camera that incorporates an Imec 25-band VNIR sensor. As depicted in Fig. 1, the camera was mounted on a sucker holder on the front hood of the vehicle and connected to a laptop inside the car through an Ethernet cable. The Photonfocus MV1 camera is a small-size snapshot camera with a GigEVision interface that can run at up to 42fps depending on its configuration. The Imec sensor is a 25-band VIS-NIR (600nm-975nm) multispectral sensor based on a CMOSIS CMV200 image wafer sensor

Fig. 1: The Photonfocus camera mounted on the front hood of the vehicle (top) and the recording laptop inside the car (bottom).

with 5µmx5µm pixel size and 2048x1088 resolution. The spectral bands are obtained by a mosaic of Fabri-Perot filters that produce 2D images with 5x5 pixel windows.

The selected optics was an Edmund Optics 16mm C Series VIS-NIR fixed focal length lens. Since the sensor dimension is 11.2 mm x 5.8 mm (2/3" format), this lens gives us a 30.9º FOV. To maximize depth of field and avoid excessive vignetting we have set the aperture to f/8. Since the Photonfocus camera does not adjust the exposure time automatically, recordings have been done by setting two different exposure times depending on light conditions: 10ms for bright light conditions and 20ms for dull days and early morning/sunset recordings. No longer exposure times have been used due to the appearance of phantom effects in fast-moving objects. A 12bit resolution has been used for raw binary information coding, while the camera throughput has been limited to 11fps to avoid excessive memory consumption.

B. Raw Image Processing

HSI snapshot mosaic cameras produce 2D gray-level images that must be transformed into hyperspectral cubes through a sequence of image preprocessing stages. The applied preprocessing pipeline comprises raw image cropping, reflectance calculation, band extraction, spatial filtering, band alignment and band normalization.

After the cropping and framing of the raw image, a reflectance signal is computed from the captured radiance values for a reliable comparison of the images' spectra. The reflectance is calculated as a normalized radiance, taking a white reference frame that is assumed to represent the maximal response. Besides the white balancing, a bias correction to eliminate static noise is carried out by previously subtracting a dark reference frame from both the image frame and the white reference frame used for the normalization:

$$\rho = \frac{target(\tau_1) - dark\_ref(\tau_1)}{white\_ref(\tau_0) - dark\_ref(\tau_0)} \frac{\tau_1}{\tau_0} \quad (1)$$

In this stage, the target image is split into 25 images with spatial resolution 409x216. Next, a median filter using a window size of 3x3 pixels is applied to every band frame. We have included this optional filtering stage since we have observed a definite positive effect on spectral separability and pixel classification indexes.

After the preprocessing, the presence of the filter mosaic pattern is removed (demosaicing) by a processing that includes a band extraction step followed by a band alignment (translation to center) performed by bilinear interpolation. Finally, it must be mentioned that we have not applied a spectral correction stage to the processing pipeline as we have not observed any improvement in the separability of the spectral signatures.

C. Itineraries and Dataset Organization

One of the main objectives of the creation of this database was to provide images of the maximum diversity regarding types of roads, light conditions and weather conditions. Thus, the database has been structured according to four main parameters: season of the year, time of day, weather conditions and type of road. The dataset contains videos recorded while driving along three road types: urban streets and roads, country and interurban roads, and highways. Driving outings have been scheduled for the four year seasons of the year and for three different times of day -dawn, full daylight and sunset- and under four weather conditions: sunny, cloudy, rainy/wet, and foggy. Images under heavy rain, ice or snow conditions have not been yet included.

Each image in the database is linked to four files:

- The binary raw file obtained from the camera (.bin).
- A Matlab Level 5 file containing a three-dimensional matrix for a 25- band hyperspectral cube obtained from the raw file (.mat).
- A Portable Networks Graphics (PNG) file containing a false color RGB image (.png).
- A PNG file containing the annotated image mask or ground-truth image (.png).

In addition, the dataset includes tarball (.tar) files containing raw video sequences of approximately 20 seconds from which the annotated images were extracted.

D. Image Annotation

This first version of the dataset (HSI-Drive v1.0) contains 276 annotated images from recordings taken during spring and summer. The total count of labeled pixels is 16,825,858. Version v1.1, incorporating images taken during fall and winter, will be released by the end of 2021. This dataset is aimed at the development of detection systems that directly rely on the separability of the spectral signature of materials and the features derived from spectral information, thus the labeling for the image annotation has been performed according to material surface reflectances as follows:

Fig. 2: False RGB image example (top) and ground-truth (bottom)

- Class1 (1): Road
- Class2 (2): Road marks
- Class3 (3): Vegetation
- Class4 (4): Painted Metal
- Class5 (5): Sky
- Class6 (6): Concrete/Stone/Brick
- Class7 (7): Pedestrian/Cyclist
- Class8 (8): Water
- Class9 (9): Unpainted metal
- Class10 (10): Glass/Transparent plastic

As a result, in an eventual HSI-supported ADS, the detection of painted metal surfaces should focus the vision system on vehicles, road signs, traffic light poles etc. Unpainted metal detection should focus systems on guardrails, metallic fences, lighting poles etc. Per-pixel image annotation has been performed manually using simple polygon labelling tools. The annotation procedure has been very conservative, selecting only the areas that clearly belong to each class, and leaving the edges and large areas of the background unlabeled, as illustrated in Fig. 2 . This approach is aimed at maximizing ML algorithms based on spectral features to the detriment of techniques that rely on spatial features. However, it is planned that future new versions of the dataset will also include dense semantic annotation files.

IV. ANALYSIS OF SEPARABILITY

Spectral separability indexes provide information about how well a classification system could potentially differentiate the hyperspectral signature of the ten different item categories or classes used for annotation. We have selected the 121 images of the dataset recorded in spring for this analysis. The distribution of labeled pixels per class contained in this data subset is shown in Table I. Due to an insufficient amount of data pertaining to Class8 (water), this class was removed.

Various criteria to evaluate the separability of classes can be found in the literature. For remote-sensing applications, in particular, the Transformed Divergence and the JeffreysMatusita distance [21] are the most used metrics. We computed both indexes for the spring subset and verified the correlation of obtained values. Since the JM index estimates the probability of correct classification, in the following we will refer to this metric. Given 2 classes i, j the distance JM is defined by the following equation:

TABLE I: Number of labeled pixels in the experimental image subset

label	1	2	3	4	5
#pixels	3,482,617	174,315	1,330,837	91,002	247,256
label	6	7	8	9	10
#pixels	383,955	30,162	1,327	29,495	32,134

$$JM_{i,j} = [2(1 - e^{-B_{i,j}}]^{1/2} \quad (2)$$

where B_{ij} is the Bhattacharyya distance which is defined by

$$B_{i,j} = \frac{1}{8}(\mu_i - \mu_j)^T \frac{\Sigma^{-1}}{2}(\mu_i - \mu_j) + \frac{1}{2} ln \frac{|\Sigma|/2}{\sqrt{|\Sigma_i||\Sigma_j|}} \quad (3)$$

with μ_i and μ_j, and Σ_i and Σ_j being the mean vectors and the covariance matrices of classes i and j respectively, and $\Sigma = \Sigma_i + \Sigma_j$

Th JM index is bounded between 0, complete overlapping of classes, and 2, complete separability. More specifically, a value between 0 and 1 indicates very poor separability; a value between 1.0 and 1.9 means moderate separability (i.e., the two signatures are separable, to some extent) and a value between 1.9 and 2.0 implies good separability.

Table II shows the JM distance e for every pair of classes. The last row shows the mean value for each class. Thus, for example, if we pay attention to the mean values, classes 5 (Sky) and 1 (Road) are the ones with the best separability from the rest of classes, showing values close to 2.0. In contrast classes 9 (Unpainted metal) and 4 (Painted Metal) are the most overlapped between them and with the rest of classes in average. The rest of the classes present intermediate values.

V. CLASSIFICATION EXPERIMENTS

Some basic pixelwise classification experiments were carried out on 62 images randomly selected form the experimental spring subset (3,778,485 labeled pixels in total). Only original spectral signatures were used as inputs and no band

TABLE II: JM distance for each pair of classes

	1	2	3	4	5	6	7	9	10
1		1.95	1.96	1.91	2.00	1.45	1.84	1.93	1.71
2	1.95		1.88	1.27	1.93	1.77	1.80	1.64	1.73
3	1.96	1.88		1.58	2.00	1.72	1.35	1.64	1.78
4	1.91	1.27	1.58		1.95	1.36	1.28	0.90	1.23
5	2.00	1.93	2.00	1.95		1.99	2.00	1.99	1.97
6	1.45	1.77	1.72	1.36	1.99		1.36	1.28	1.13
7	1.84	1.80	1.35	1.28	2.00	1.36		1.32	1.43
9	1.93	1.64	1.64	0.90	1.99	1.28	1.32		1.26
10	1.71	1.73	1.78	1.23	1.97	1.13	1.43	1.26	
Mean	1.84	1.75	1.74	1.44	1.98	1.51	1.55	1.50	1.53

TABLE III: Architecture and performance evaluation of the ANN classifiers: model hyperparameters, total number of adjustable parameters, and test accuracy figures for each experiment (Acc.0 applies to the "other" class)

	Exp.1	Exp.2	Exp.3	Exp.4
ANN	25-25-300-3	25-25-300-4	25-25-300-5	25-25-300-9
#params	9.4K	9.7K	10K	11.2K
Acc.1 %	95.83	94.34	93.37	91.71
Acc.2 %	93.86	91.49	90.56	87.64
Acc.3 %	NA	NA	NA	94.37
Acc.4 %	NA	91.16	88.87	84.13
Acc.5 %	NA	NA	NA	98.81
Acc.6 %	NA	NA	NA	84.14
Acc.7 %	NA	NA	84.91	82.48
Acc.9 %	NA	NA	NA	72.08
Acc.10 %	NA	NA	NA	73.34
Acc.0 %	94.10	90.28	88.04	NA
OA %	95.15	92.79	91.36	92.16
AA %	94.60	91.82	89.16	85.41

selection neither feature extraction procedures were applied. Obviously, segmentation results can be eventually improved by using the more sophisticated machine learning schemes that incorporate combined spectral-spatial feature extraction techniques or by essaying deep learning models. However, such analysis is beyond the scope of this paper and will be addressed in future publications. The intention of this preliminary experimentation was to show the discrimination capacity of baseline ML models such as simple shallow ANN (Artificial Neural Networks) with low computational cost. With this aim, in the following subsections we show some results obtained from four different classification experiments performed when training shallow ANN classifiers.

A. Experimental setup

After some preliminary analysis, an ANN topology with two hidden layers (25-L1-L2-m) was selected as the base model for the classifiers, where *L1* and *L2* are the number of neurons in the first and second hidden layers respectively and *m* is the number of different classes to be categorized at each experiment. For each experiment, several networks were trained by means of the Levenberg-Marquardt backpropagation algorithm to maximize the classification's overall accuracy (OA) in the validation set. Both OA and the average accuracy (AA) of pixel classification were adopted as performance metrics. Training sets were created including randomly selected 50,000 pixels, when available, from each class selected for classification at each experiment. Of these, 10,000 pixels were used for the validation set to prevent overfitting. When insufficient data were available for a certain class, the corresponding training subset was reduced: 25,000 pixels for Class7, and 12,000 pixels for both class9 and Class10. This means that, out of the 3,778,485 pixels in the experimental dataset, a maximum of 340,000 were used for training (10%), while the rest of available data were used for testing: this is quite a challenging setup.

B. Experiment 1: drivable/no-drivable

In the first experiment the labels were grouped to train the ANNs with only three classes: "road", "road marks" and "other". This system acts therefore as a sort of drivable region detector that could be enhanced with lane departure and trajectory generator capabilities. The training set for this experiment contained 150,000 out of 3,778,485 pixels, i.e., only 4% of available data. The obtained classification figures for this and the subsequent experiments are summed up in table III. As expected from the separability measures shown in Table II, high classification accuracy is achieved for both Class1 (road) and Class2 (road marks). Figure 4 shows the pixelwise segmentation of three example images for highway, road and urban scenes. To enhance image segmentation a two-stage spatial regularization (SR) algorithm was applied to the ANN output: first, the pixels classified with a low confidence (ANN output values below 0.8) were re-labeled as "don't know", and then the SR process assigned a new label to those pixels by a majority voting criterion over 5x5 pixel windows. As can be seen, although the obtained results are generally correct, the vehicles on the road and urban scenes present many misclassified pixels belonging to the "road marks" category. This is due to the highly reflective surfaces of the car bodies painted in white.

C. Experiment 2: drivable/road signals-vehicles

In the second experiment the target categories to be detected were "road", "road-marks", "painted-metal" and "other". This system should provide information to perform not only as a lane tracking system, but also to focus the system on image areas that potentially contain vehicles, road signals and other painted metallic surfaces. The training set for this experiment contained 200,000 pixels, i.e. 5.3% of available data. In the example highway and road test scenes, the system correctly detects road signals and the presence of vehicles, although there are some misclassified pixels in the background (see Fig.5). However, in the more complex urban scene practically all the background (except the sky) has been classified as painted metal. This is not surprising since, according to the figures in Table II, the separability of class4 with some of the rest of the classes is quite poor.

D. Experiment 3: drivable/road signals-vehicles/pedestrians

In this experiment, the pixels labeled as "pedestrian/cyclist" are incorporated into the training set as a separate class. Thus, the resulting ADS could add a focus on any people at sight to the capacities described in Experiment 2. The training set for this experiment contained 225,000 pixels i.e. 6% of available data. The obtained classification figures are summed up in Table III. Accuracy values are not so good for Class4 and Class7. These results are consistent, indeed, with the figures in the separability tables. As can be observed in the sample images in Fig. 6, highway and road scenes show some additional false positives for the "pedestrian". In the urban scene, pedestrians are quite efficiently segmented, but here again the segmentation of the image background is faulty.

E. Experiment 4: all classes

The last experiment explored the potential of these simple classifiers to produce a complete segmentation of the images

comprising all categories used in the labeled image dataset (except for the "water" class, as explained above). As shown in Table III, there are some classes with testing accuracies of over 90% ("road","vegetation", and "sky") while the classes "road marks", "painted metal", "concrete-stone" and "pedestrian" show accuracy figures under 90%. Two classes, "unpainted metal" and "glass/transparent plastic", have accuracy figures under 80%. However, it must be taken into account that these are precisely the two classes that contribute the least amount of data to the training set. In Fig. 7 we see qualitatively quite acceptable image segmentation for the highway and road scenes. In the urban scenes, road, road marks and items in the foreground are quite correctly detected. Nevertheless, the buildings in the background are misclassified, although the sky and the receding vehicle are correctly detected.

VI. CONCLUDING REMARKS

The use of HSI sensors in ADS is expected to grow substantially in the following years. The incorporation of hyperspectral information will make it possible to improve the accuracy and robustness of these systems as well as, eventually, reduce the computational burden of current image processing pipelines. To meet these objectives it will be necessary, however, to deepen the research on the HSI processing applied to ADS. The HSI-Drive dataset has been created with the desire to contribute to the research in this field. Unlike other recently reported similar datasets, HSI-drive has been designed to provide a structured image dataset of wide diversity in terms of driving scenarios, lighting conditions, and weather conditions. Current annotation of images has been focused on the spectral reflectance characteristics of various material surfaces relevant to the development of ADS. In this work, we show that even a simple perpixel processing of pure spectral information obtained in the NIR spectrum can produce quite accurate image segmentation with baseline neural classifiers. The use of more sophisticated algorithms that incorporate feature extraction stages, together with spatial/contextual information should enhance system performance. The HSI-drive dataset will evolve in the years to come. A larger version of the dataset that will incorporate new hyperspectral videos and annotated images corresponding to the winter and fall seasons is currently under development and is expected to be ready by the end of the year. We also plan to add dense semantic labeling of images in the future.

ACKNOWLEDGMENT

Thanks to Zorion Doistua for his selfless help acting as driver of our vehicle for so many outings, and many thanks to Anton Basterretxea for his willingness to help as a recording operator when requested.

REFERENCES

[1] D. Feng, C. Haase-Shütz, L. Rosenbaum, H. Hertlein, C. Gläser, F. Timm, W. Wiesbeck, and K. Dietmayer, "Deep Multi-Modal Object Detection and Semantic Segmentation for Autonomous Driving: Datasets, Methods, and Challenges," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1341-1360, March 2021.

[2] E. Yurtsever, J. Lambert, A. Carballo and K. Takeda, "A Survey of Autonomous Driving: Common Practices and Emerging Technologies," *IEEE Access*, vol. 8, pp. 58443-58469, March 2020.

[3] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern Trends in Hyperspectral Image Analysis: A Review," *IEEE Access*, vol. 6, pp. 14118-14129, March 2018.

[4] M. West, J. Grossmann, and C. Galvan, "Comercial Snapshot Spectral Imaging: The Art of the Possible", Mitre Technical report, Mitre Corporation, September 2018.

[5] P. Govardhan and U. C. Pati, " NIR image based pedestrian detection in night vision with cascade classification and validation," in *Proc. Int. Conf. on Advanced Communication Control and Computing Technologies*, Tamilnadu, India, 2014, pp. 1435-1438.

[6] M. Velte, "Semantic image segmentation combining visible and near infrared channels with depth information." Ph.D. dissertation, Bonn-Rhein-Sieg University of Applied Sciences, 2015.

[7] E. S. Jeon, J.-S. Choi, J. H. Lee, K. Y. Shin, Y. G. Kim, T. T. Le, and K. R. Park, "Human detection based on the generation of a background image by using a far-infrared light camera," *Sensors* vol. 15, no. 3, pp.6763–6788, March 2015.

[8] A. González, Z. Fang, Y. Socarras, J. Serrat, D. Vázquez, J. Xu, and A. M. López, "Pedestrian detection at day/night time with visible and FIR cameras: A comparison," *Sensors*, vol. 16, no. 6, pp. 820, June 2016.

[9] S. Hwang, J. Park, N. Kim, Y. Choi, and I. S. Kweon, "Multispectral pedestrian detection: Benchmark dataset and baseline," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Boston, MA, USA, 2015, pp. 1037-1045.

[10] J. Wagner, V.r Fischer, M. Herman, and S. Behnke, "Multispectral pedestrian detection using deep fusion convolutional neural networks," in *Proc. European Sympousium on Artificial Neural Networks*, Bruges, Belgium, 2016.

[11] T. Karawasa, K. Watanabe, Q. Ha, A. Tejero-De-Pablos, Y. Ushiku, and T. Harada, "Multispectral Object Detection for Autonomous Vehicles," in *Thematic Workshops'17*, Mountain View, CA, USA, October 2017.

[12] Q. Ha, K. Watanabe, T. Karasawa, Y. Ushiku and T. Harada, "MFNet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, Vancouver, BC, Sept. 2017, pp. 5108-5115.

[13] J. Herweg, J. Kerekes, and M. Eismann, "Separability between pedestrians in Hyperspectral Imagery," *Applied Optics*, vol. 52, no. 6, pp. 1330-1338, 2013.

[14] C. Winkens, F. Sattler, P. Dietrich, "Hyperspectral Terrain Classification for Ground Vehicles," in *Proc.12th Int. Joint Conf. Computer Vision, Imaging and Computer Graphics Theory and Applications*, Porto, Portugal, 2017, pp. 417-424.

[15] C. Winkens, V. Kobelt, D. Paulus, "Robust Features for Snapshot Hyperspectral Terrain- Classification," in *Proc. 17th Int. Conf. Computer Analysis of Images and Patterns*, Ystad, Sweden, August 2017, pp. 16-27.

[16] C. Winkens, D. Paulus, "Context aware hyperspectral scene analysis," in *Proc. Int. Symp. Electronic Imaging Science and Technology*, Springfield, USA, 2018, pp. 346.1-346.7.

[17] Y. Huang, E. Huang, L. Chen, S. You, Y. Fu, Q. Shen "Hyperspectral Image Semantic Segmentation in Cityscapes," arXiv preprint, arXiv:2012.10122v1, Dec. 2020.

[18] C. Winkens, F. Sattler, V. Adams and D. Paulus, "HyKo: A Spectral Dataset for Scene Understanding," in *Proc. IEEE Int. Conf. Computer Vision Workshops*, Venice, Italy, 2017, pp. 254-261.

[19] S. You, E. Huang, S. Liang, Y. Zheng, Y. Li, F. Wang, S. Lin, Q. Shen, X. Cao, D. Zhang, "Hyperspectral city v1.0 dataset and benchmark," arXiv preprint arXiv:1907.10270, Feb. 2020.

[20] J. Lu, H. Liu, Y. Yao, S. Tao, Z. Tang and J. Lu, "Hsi Road: A Hyper Spectral Image Dataset For Road Segmentation," in *Proc. IEEE Int. Conf. Multimedia and Expo (ICME)*, London, UK, July 2020, pp. 1-6.

[21] G.Forestier, J. Inglada, C. Wemmert, and P. Gancarski, "Comparison of optical sensors discrimination ability using spectral libraries," *International Journal of Remote Sensing*, vol. 34, no. 7, pp.2327-2349, 2013.

Fig. 3: False color example images generated from HSI cubes: a highway scene (left), a road scene (centre), and an urban scene (right)

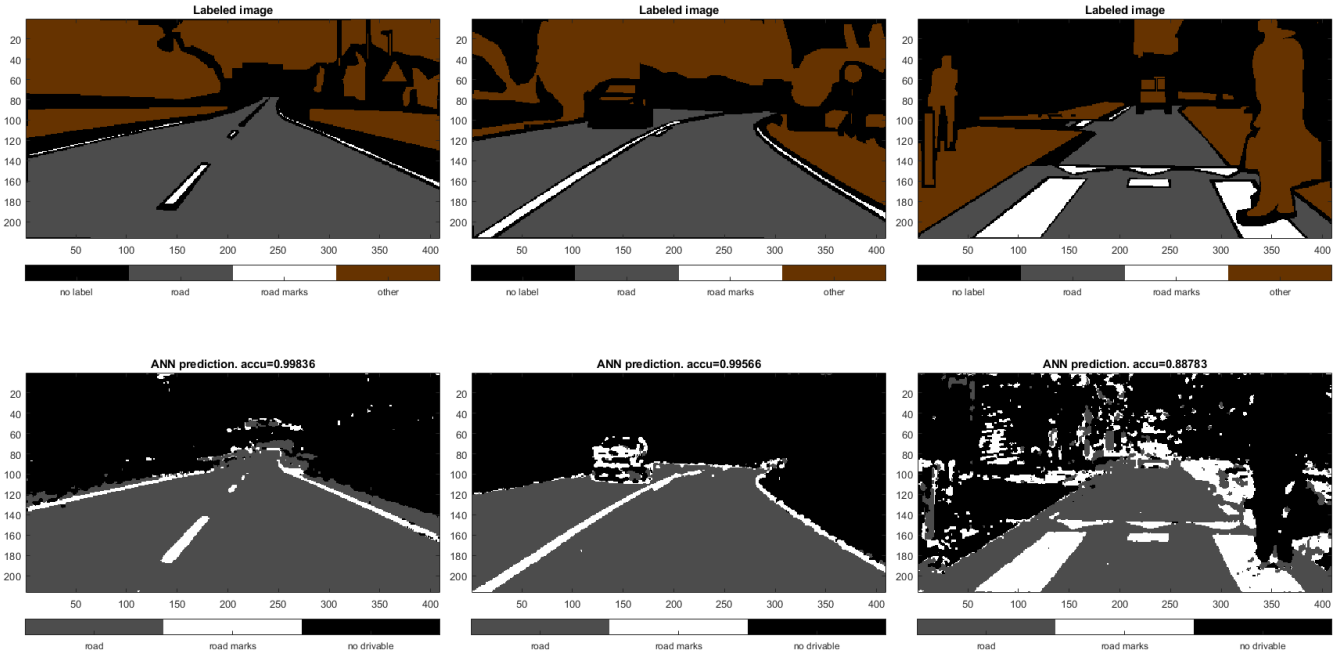

Fig. 4: Image segmentation examples for experiment 1: ground-truth images (top) and segmented images (bottom)

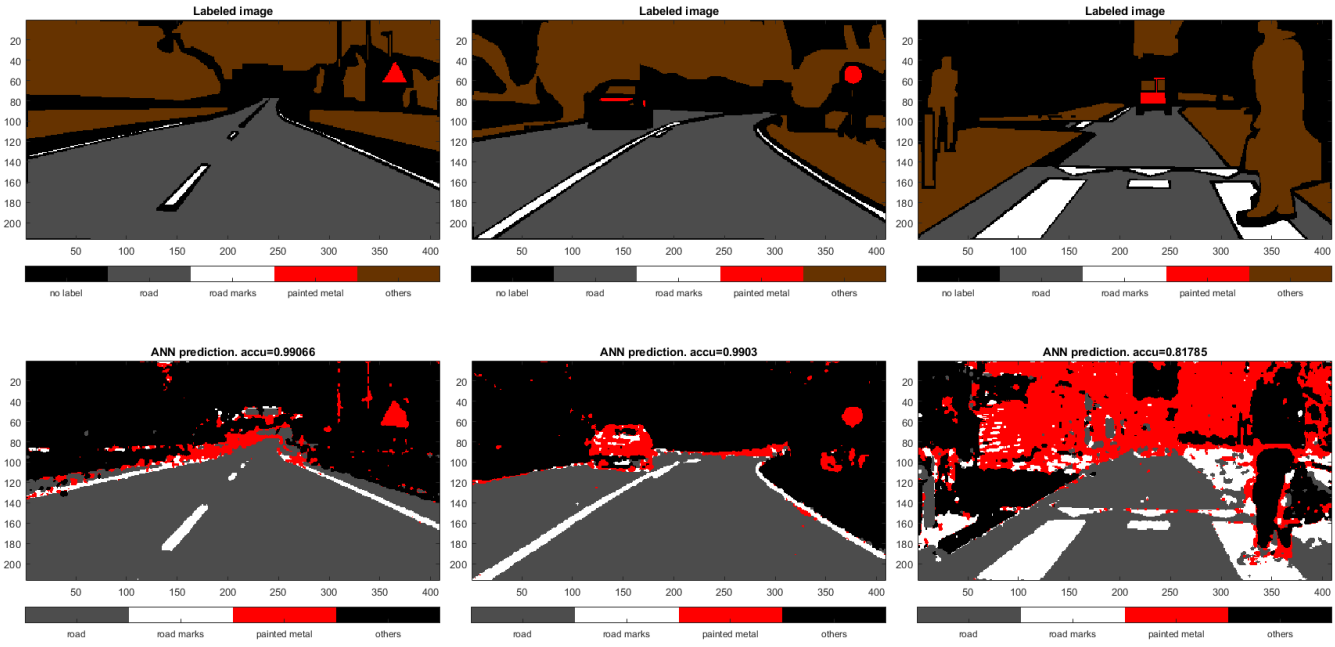

Fig. 5: Image segmentation examples for experiment 2: ground-truth images (top) and segmented images (bottom)

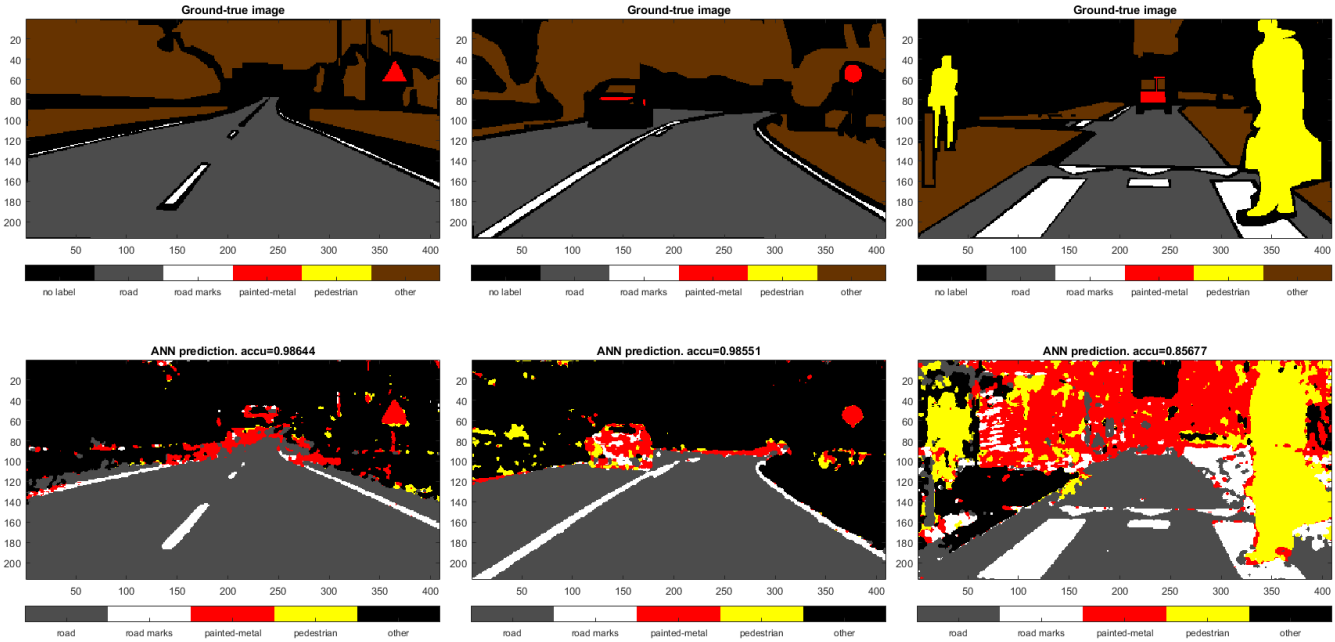

Fig. 6: Image segmentation examples for experiment 3: ground-truth images (top) and segmented images (bottom)

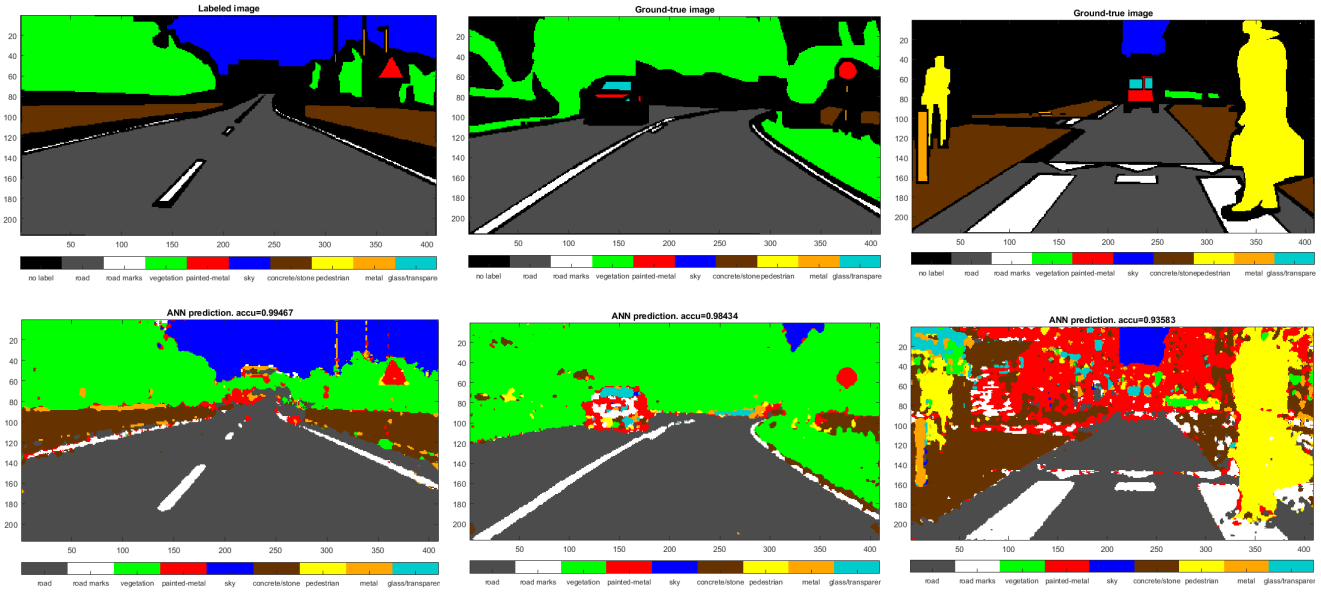

Fig. 7: Image segmentation examples for experiment 4: ground-truth images (top) and segmented images (bottom)