

HyKo: A Spectral Dataset for Scene Understanding

Christian Winkens, Florian Sattler, Veronika Adams, Dietrich Paulus
University of Koblenz-Landau

{cwinkens, sflorian92, vadams, paulus}@uni-koblenz.de

Abstract

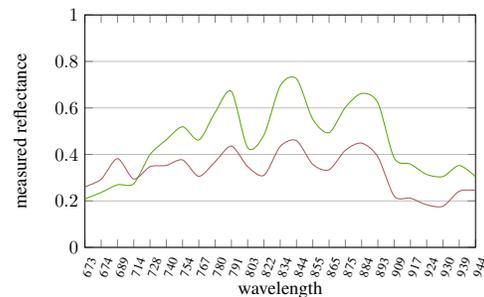
We present datasets containing urban traffic and rural road scenes recorded using hyperspectral snap-shot sensors mounted on a moving car. The novel hyperspectral cameras used can capture whole spectral cubes at up to 15 Hz. This emerging new sensor modality enables hyperspectral scene analysis for autonomous driving tasks. Up to the best of the author's knowledge no such dataset has been published so far.

The datasets contain synchronized 3-D laser, spectrometer and hyperspectral data. Dense ground truth annotations are provided as semantic labels, material and traversability. The hyperspectral data ranges from visible to near-infrared wavelengths. We explain our recoding platform and method, the associated data format along with a code library for easy data consumption. The datasets are publicly available for download.

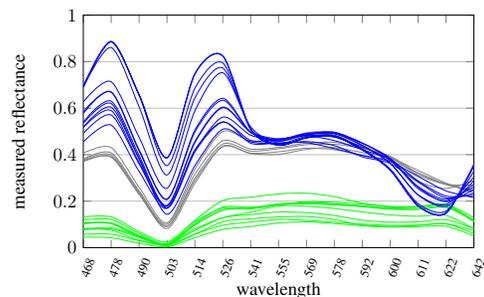
1. Introduction

In the field of autonomous driving environment perception and, in particular, drivability analysis is a key factor. Given input data, the correct semantic interpretation of a road scene is crucial for successfully navigating autonomous cars. Often laser scanners are used for environmental perception where the geometric nature of the environment or the difference in height of individual segments are investigated to classify the scene.

On closer examination of the established methods, it is quickly clear that the complexity of traffic scenes and especially unstructured environments in off-road scenarios can not be adequately depicted and recorded utilizing only unimodal sensory. In order to achieve a comprehensive representation of the environment, the use of multimodal sensory is necessary, which may be achieved by the additional use of suitable camera systems. Here the hyperspectral imaging could provide great advantages. Hyperspectral data allow a more detailed view of the composition and texture of materials, plants and floor coverings than normal cameras. It is an important and fast growing topic in remote



(a) Example plots of hyperspectral spectra captured with the NIR-camera. Street (gray) and vegetation (green).



(b) Example plots of hyperspectral spectra captured with the VIS-camera. Sky (blue) street (gray) and vegetation (green).

Figure 1: Examples of spectral plots of NIR-camera and VIS-camera for different materials.

sensing where typically a few tens to several hundreds of contiguous spectral bands are captured. Researchers use so called line-scanning sensors on Landsat, SPOT satellites or the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) for acquiring precise spectral data. These line-scanning sensors provide static information of the Earth's surface and allow a static analysis. This area has been firmly established for several years and is essential for many applications like earth observation, inspection and agriculture. Additionally onboard realtime hyperspectral image analysis for autonomous navigation is an exciting and promising new application scenario. We want to explore this new research area and investigate the use of hyperspectral imaging

in the area of autonomous driving in traffic and off-road scenarios. But a drawback of established sensory are the scanning requirements for constructing a full 3-D hypercube of a scene. By utilizing line-scan cameras, multiple lines need to be scanned, while for cameras using special filters several frames have to be captured to construct a spectral image of the scene. The slow acquisition time is responsible for motion artifacts which impede the observation of dynamic scenes. Therefore, new sensor techniques and procedures are needed here.

This drawback can be overcome with novel highly compact, low-cost, snapshot mosaic (SSM) imaging using the Fabry-Perot principle. Since this technology can be built in small cameras and the capture time is considerably shorter than that of filter wheel solutions allowing to capture a hyperspectral cube at one discrete point in time. We used this new sensory class and installed these hyperspectral cameras together with additional sensory on land vehicles driving through urban and suburban areas building an all new dataset for semantic scene analysis.

The main purpose of this dataset is to investigate the use of hyperspectral data for computer vision especially in road scene understanding to potentially enhance drivability analysis and autonomous driving even in off-road scenarios. An example of several spectras of classes we reconstructed is given in Figure 1, where differences in the spectral reflectances of different material classes can be spotted.

In this paper we present new datasets including raw data and several hundred hand labeled hyperspectral data cubes (hypercubes) captured with the use of snapshot mosaic hyperspectral cameras which were mounted together with additional sensory on land vehicles. We want to investigate the use of hyperspectral data for scene understanding in autonomous driving scenarios. The remainder of this paper is organized as follows. In the following section 2 an overview of already available datasets for autonomous driving and especially hyperspectral datasets is given. Then our general setup is presented in section 3. The preprocessing and dataset format is explained in section 4. The dataset itself and annotations are described in section 5. Finally a conclusion of our work is given in section 6.

2. Related Work

Hyperspectral image classification has been under active development recently. Given hyperspectral data, the goal of classification is to assign a unique label to each pixel vector so that it is well-defined by a given class. The availability of labeled data is quite important for nearly all classification techniques. While recording data is usually quite straightforward, the precise and correct annotation of the data is very time-consuming and complicated.

There is only little research in literature on hyperspectral classification utilizing terrestrial spectral imaging, where

data was not captured from an earth orbit or an airplane but from cameras which were mounted on land-based vehicles. One of the few examples is the vegetation detection in hyperspectral images as demonstrated by Bradley *et al.* [2], who showed that the use of the Normalized Difference Vegetation Index (NDVI) can improve classification accuracy. Additionally Namin *et al.* [11] proposed an automatic system for material classification in natural environments by using multi-spectral images consisting of six visual and one near infrared band.

In the field of classical computer vision, there are now some established data sets. For example, Geiger *et al.* [10] presented a dataset of several hours using a variety of sensor modalities like color cameras, captured from a car for use in autonomous driving and robotics research. The captured scenarios vary from inner-city traffic scenes to rural areas. Another benchmark dataset consisting of 500 stereo image pairs was published by Scharwchter *et al.* [13]. They provide ground truth segmentations containing five object classes. Blanco *et al.* [1] published six outdoor datasets including laser and color camera data with an accurate ground truth. Skauli *et al.* [14] published a dataset of hyperspectral images displaying faces, landscapes and buildings which cover wavelengths from 400 – 2000nm. Which extends the hyperspectral face database, captured in the visible light spectrum and published by Wei Di *et al.* [5].

Yasuma *et al.* [19] presented a framework for designing cameras that can simultaneously capture extended dynamic range and higher spectral resolutions. They captured 31 hyperspectral images (400 – 700nm) of several static scenes, with a wide variety of materials, using a tunable filter and a cooled CCD camera.

Estimates of the frequency of metameric surfaces, which appear the same to the eye under one illuminant but different under another, were obtained from 50 hyperspectral images of natural scenes [7]. Foster *et al.* [6] published a small dataset where they captured time-lapse hyperspectral radiance images five outdoor vegetated and non-vegetated static scenes. They explored illumination variations on color measurements. A collection of fifty hyperspectral images of indoor and outdoor scenes captured under daylight illumination and twenty-five images captured under artificial and mixed illumination was published by Chakrabarti and Zickler [4]. The problem of segmenting materials in hyperspectral images was recently investigated by Yu *et al.* [20]. They use a per pixel classification followed by a super-pixel based relabelling to classify materials. For evaluation they collected 51 portrait and landscape hyperspectral images from several above mentioned databases [7, 19, 4, 14] which have been processed to obtain a common sampling in the range of 430 – 700nm. To the best of our knowledge there are only two other datasets which use the same sensor as we did. Fotiadou *et al.* [8] propose a deep feature learning ar-

chitecture for efficient feature extraction, encoding spectral-spatial information of hyperspectral scenes. They focused on the classification of snap-shot mosaic hyperspectral images and built a new hyperspectral classification dataset of indoor scenes consisting of 90 images. The images were acquired under different illuminations and different view-points displaying 10 different objects like bananas, glasses and other things. The combination of RGB and hyperspectral data was evaluated by Cavigelli *et al.* [3] on data with static background and a very small dataset with only a few images utilizing deep neural nets.

As far as we know, there is no publicly available dataset with hand labeled hyperspectral data recorded by snapshot mosaic hyperspectral cameras mounted on a moving land vehicle capturing driving scenarios. It's our goal to provide new datasets which help to investigate the use of hyperspectral data in driving scenarios.

3. Sensor Setup

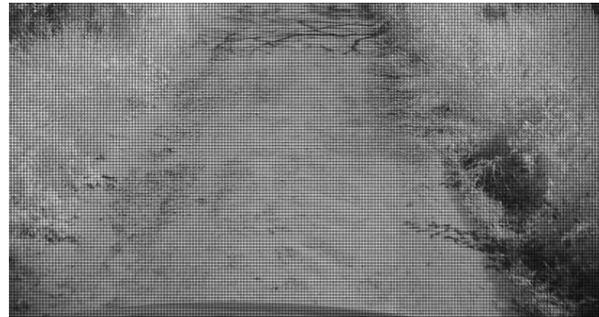
We have mainly recorded two datasets, called *HyKo1* and *HyKo2* with different sensor setups which will be illustrated and described in the next sections. As hyperspectral camera sensors we used the MQ022HG-IM-SM4X4-VIS (*VIS*) and MQ022HG-IM-SM5X5-NIR (*NIR*) manufactured by Ximea with an image chip from IMEC [9] utilizing a snapshot mosaic filter which has a per-pixel design. The filters are arranged in a rectangular mosaic pattern of n rows and m columns, which is repeated w times over the width and h times over the height of the sensor. These sensors are designed to work in a specific spectral range which is called the active range which is:

- visual spectrum (*VIS*): 470–630 nm
- near infrared spectrum (*NIR*): 600-975 nm

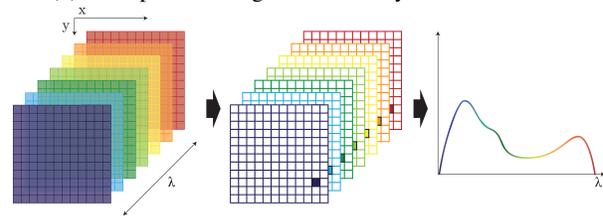
The *VIS* camera has a 4×4 mosaic pattern and the *NIR* 5×5 which results in a spatial resolution of approx. 512×272 pixels (4×4) and 409×217 pixels (5×5). Ideally every filter has peaks centered around a defined wavelength spectrum with no response outside. However contamination is introduced into the response curve and the signal due to physical constraints. These effects can be summarized as a spectral shift, spectral leaking, and crosstalk and need to be compensated [16, 18, 15, 17]. Additionally we used a acA2040-25gc color camera from Basler, with a LM6HC lens from Kowa in the *HyKo1* dataset. So it's possible to examine differences between color and hyperspectral data. The hyperspectral cameras and the color camera are synchronized using a custom build hardware trigger. Unfortunately we couldn't synchronize the Velodyne HDL-32E along with cameras, because there is no direct interface for triggering the Velodyne HDL-32E itself. For the *VIS* and the *NIR* camera we used the LM5JC10M lens from Kowa.



(a) Example raw image data taken by the *VIS* camera.



(b) Example raw image data taken by the *NIR* camera.



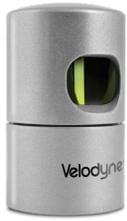
(c) A schematic representation of a hypercube and an interpolated plot of a single data point.

Figure 2: Raw image *VIS* camera with visible mosaic pattern. And a schematic representation of a hypercube constructed from the captured data.



Figure 3: Hyperspectral cameras used to capture the new datasets.

Additionally the *NIR* camera was equipped with a 16mm VIS-NIR lens from Edmund Optics in some sequences, like displayed in Figure 3. Some example raw data can be seen



(a) Velodyne HDL-32E¹



(b) Qmini Wide

Figure 4: Additional sensory utilized in our datasets.

in Figure 2.

Additionally we used several Velodyne HDL-32E sensors to gather 3D pointclouds in our datasets. The sensor rotates with 10HZ has 32 laser beams and delivers 700 000 3D points per second. It has a 360 deg horizontal +10 to -30 vertical field of view.

To get a ground truth of the illumination of the scene we integrated a Qmini Wide spectrometer manufactured by RGB Photonics with an USB interface. It delivers 2500 spectral reflectance values from 225 – 1000nm through an optical fiber element, like displayed in Figure 4.

3.1. Sensor Calibration

Sensor calibration is crucial for sensor fusion. We calibrated our cameras intrinsically before every run and did an extrinsic calibration to allow a registration and data fusion with the Velodyne HDL-32E sensors. For calibrating multiple cameras with multiple laser range finders we developed a robust and complete methodology which is publicly available and will be published anytime soon. Here we will give a brief introduction. Registering laserscanners with cameras is non-trivial as correspondences are hard to establish, known 3D reference geometry is required. Therefore our system consists of two phases. First we perform bundle adjustment² on a custom build marker system to obtain a reference 3D geometry using camera images. Then we observe this geometry in different poses with all the sensors that need calibration and perform non-linear optimization³ to estimate the extrinsic transformation. It allows the accurate estimation of the 3D poses of configurations of optical markers given a set of camera images.

4. Data Preprocessing and Format

In this section we briefly describe our preprocessing and data format used for our datasets.

¹Source: <http://velodynelidar.com/hdl-32e.html>

²https://github.com/cfneuhaas/visual_marker_mapping

³https://github.com/cfneuhaas/multi_dof_kinematic_calibration

4.1. Preprocessing

During recording, the hyperspectral cameras provide the images in a lossless format with 8 bits per sample. Therefore the raw data captured by the camera needs a special preprocessing. We need to construct a hypercube with spectral reflectances from the raw data like seen in Figure 2. This step consists of cropping the raw-image to the valid sensor area, removing the vignette and converting to a three dimensional image, which we call a *hypercube*. Reflectance calculation is the process of extracting the reflectance signal from the captured data of an object. The purpose is to remove the influence of the sensor characteristics like quantum efficiency and the illumination source on the hyperspectral representation of objects. We define a hypercube as $\mathcal{H} : \mathcal{L}_x \times \mathcal{L}_y \times \mathcal{L}_\lambda \rightarrow \mathbb{R}$ where $\mathcal{L}_x, \mathcal{L}_y$ are the spatial domain and \mathcal{L}_λ the spectral domain of the image. A visual interpretation of such a hypercube is displayed in Figure 2c. The hypercube is understood as a volume, where each point $\mathcal{H}(x, y, \lambda)$ corresponds to a spectral reflectance. Derived from the above definition a spectrum χ at (x, y) is defined as $\mathcal{H}(x, y) = \chi$, where $\chi \in \mathbb{R}^{|\mathcal{L}_\lambda|}$ and $|\mathcal{L}_\lambda| = n \cdot m$. The image with only one wavelength, called a spectral band $\mathcal{H}(z) = \mathcal{B}_{\lambda=z}$, is defined as follows: $\mathcal{B}_\lambda : \mathcal{L}_x \times \mathcal{L}_y \rightarrow \mathbb{R}$. This image contains $\mathbf{x} = (x, y)$ the wavelength sensitivity λ for each coordinate. After our preprocessing we get the following hypercube dimensions:

VIS 510 × 254 × 16

NIR 407 × 214 × 25

4.2. Data Format

The extracted hyperspectral data is saved as MATLAB level 5 mat-files⁴. Every mat-file holds exactly one frame with the following data.

image Holds the raw image as captured by the camera like shown in Figure 2. The raw image has a resolution of 2048 × 1088pixels for both cameras. Depending on the camera type, a defined number of pixels at the borders is not valid.

data Holds the preprocessed hyperspectral data in form of a hypercube as described in section 4.1.

label_* Holds the labels, in form of a mask, which were assigned to the data during the annotation step.

wavelengths Holds a list with the primary wavelength sensitivities for every band in the hypercube as defined by the manufacturer.

⁴https://www.mathworks.com/help/pdf_doc/matlab/matfile_format.pdf

5. Dataset

Our dataset can be accessed and downloaded through <https://wp.uni-koblenz.de/hyko/>. The website provides the hyperspectral data as well as source code for loading and more detailed technical information about the data. Further details about usage rights are given on the website.

At the moment we provide raw datalogs as *rosbags*⁵ recorded using the middleware ROS[12] where each used sensor publishes data on its own topic. Additionally we provide annotated hyperspectral data-cubes which were extracted and preprocessed from the *rosbags* in the form of matlab-files like described in section 4.2. For the *HyKo1* dataset the following sensors were mounted on a car:

- One *VIS* and one *NIR* camera
- One color camera
- One Velodyne HDL-32E
- One Qmini Wide spectrometer

For the *HyKo2* dataset the following sensors were mounted on a truck:

- One *VIS* and one *NIR* camera
- Two Velodyne HDL-32E
- One Qmini Wide spectrometer

In all our datasets all cameras are synchronized using a custom build hardware trigger.

5.1. Data extraction

We took the *rosbags* and extracted a hyperspectral cube of each camera every 4 seconds, so we got several hundred hypercubes each representing one *frame*. Due to the fact that our sensors are mounted on land vehicles we have to deal with illumination changes and direct sunlight on the sensor which distorts the reflectance calculation. Two examples are given in Figure 5, where the illumination is too strong or too weak. So we filtered the data after extraction and sorted out all the data in which more than 20 percent of the data was over or underexposed. The filtered data was then used in the annotation process.

5.2. Annotations

As there is no labeling tool available, which is able to handle our hyperspectral data correctly, we developed our own, which will be made publicly available too. The annotations were done by hand for all hypercubes of each dataset. During the labeling process not all hyper-pixels have been

⁵<http://wiki.ros.org/rosbag>

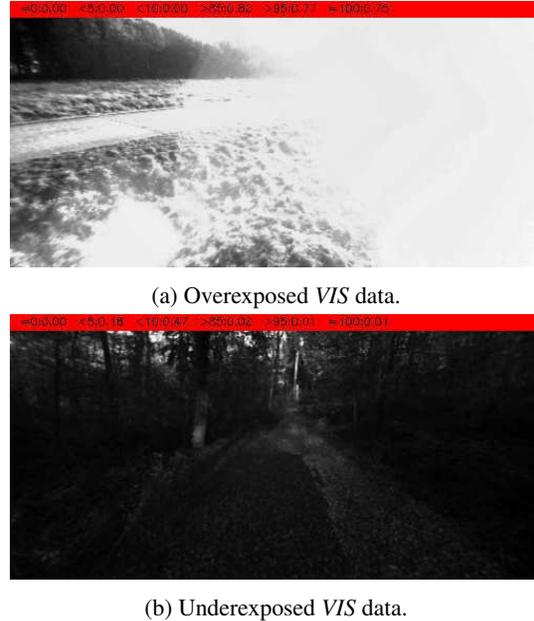


Figure 5: Examples of filtering over and underexposed data captured with the *VIS* camera from the *HyKo2* dataset.

assigned classes. This is due to the fact, that border areas between materials are not unambiguously assignable. And as later results have shown no errors arised from this constraint. For annotation we introduced three annotation classes called semantic, material and drivability, as shown in Figure 6. These annotation classes were inspired by the annotations of other data sets from the areas of hyperspectral image processing and semantic scene analysis. So over the past months we annotated several hundred hypercubes.

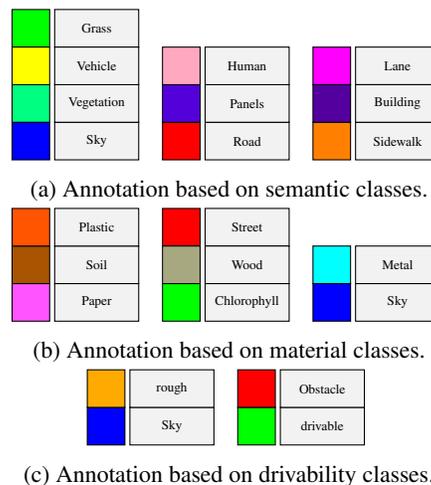
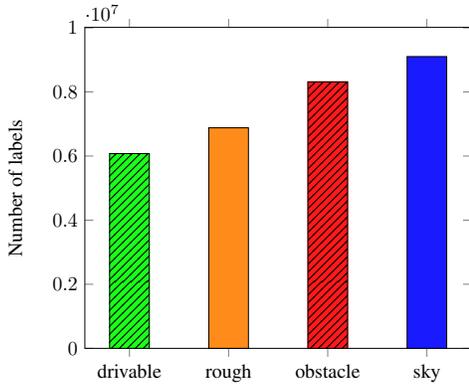
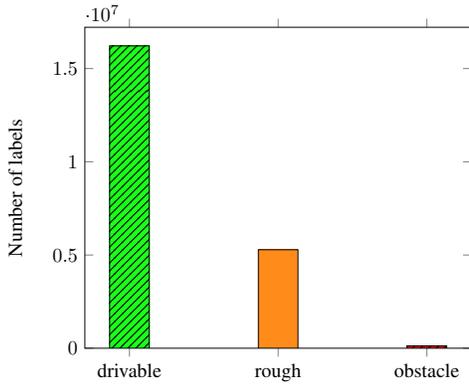


Figure 6: Introduced annotation classes

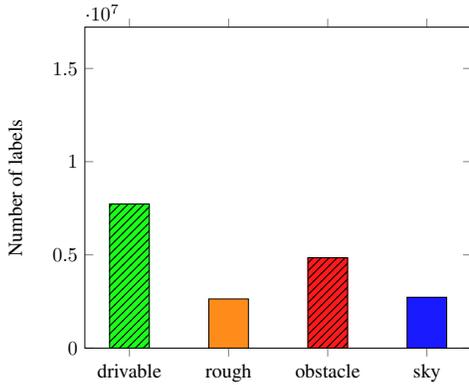
Statistics of our labeling work are displayed in Figure 7.



(a) Statistics of dataset *HyKo1* captured with *VIS* camera and drivability labels.



(b) Statistics of dataset *HyKo1* captured with *NIR* camera and drivability labels.

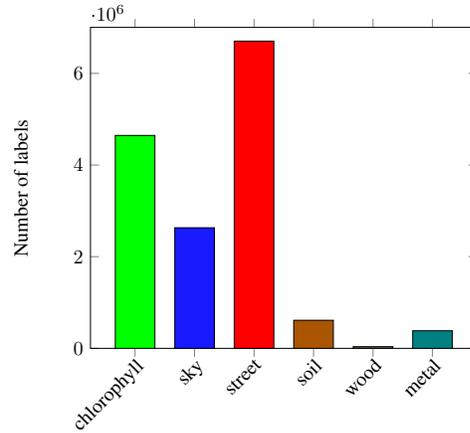


(c) Statistics of dataset *HyKo2* captured with *VIS* camera and drivability labels.

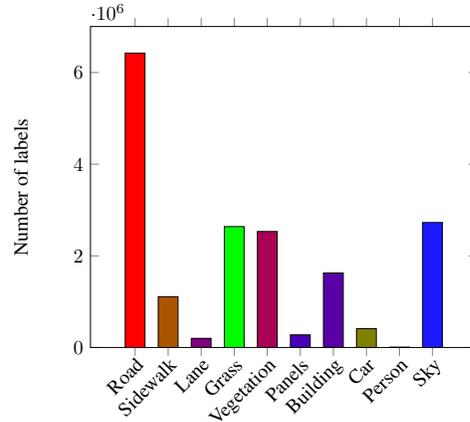
Some examples of our annotations in different annotation classes are presented in Figure 8.

6. Discussion and Conclusion

In this work, we have presented a freely available synchronized and calibrated autonomous driving dataset capturing different scenarios. It consists of hours of raw-data streams and several hundred hyperspectral data cubes. To the best



(d) Statistics of dataset *HyKo2* captured with *VIS* camera and material labels.



(e) Statistics of dataset *HyKo2* captured with *VIS* camera and semantic labels.

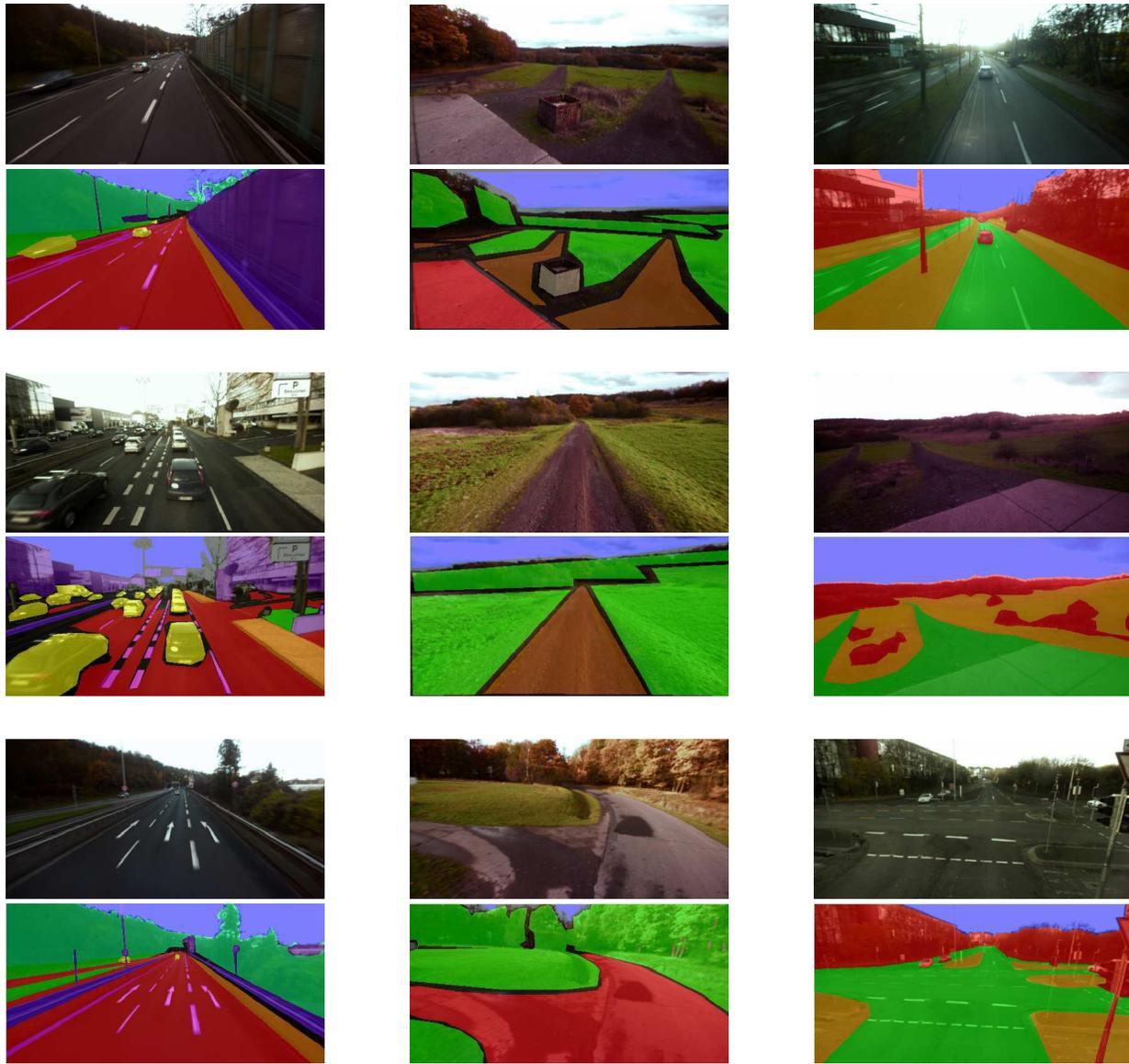
(f) Tulf Dataset vis semantic

Figure 7: Object and label occurrence statistics of our dataset.

of our knowledge it's the first dataset including snapshot mosaic hyperspectral hyperspectral data from the visible to the near infrared range, color image data, 3D pointclouds and illumination information captured by a spectrometer. We provide semantic, material and drivability labels for many of the sequences. Our goal is to examine the use of hyperspectral data for semantic scene understanding especially in autonomous driving scenarios. Furthermore we are constantly working on improving our dataset and we will continue integrating new scenes and labeled data into our datasets.

References

- [1] J.-L. Blanco, F.-A. Moreno, and J. Gonzalez. A collection of outdoor robotic datasets with centimeter-



Semantic class annotations

Spectral class annotations.

Drivability class annotations.

Figure 8: Some Example of datasets and annotations.

accuracy ground truth. *Autonomous Robots*, 27(4):327–351, 2009. 2

- [2] D. M. Bradley, R. Unnikrishnan, and J. Bagnell. Vegetation detection for driving in complex environments. In *Robotics and Automation, 2007 IEEE International Conference on*, pages 503–508. IEEE, 2007. 2
- [3] L. Cavigelli, D. Bernath, M. Magno, and L. Benini. Computationally efficient target classification in multispectral image data with deep neural networks. *arXiv preprint arXiv:1611.03130*, 2016. 3
- [4] A. Chakrabarti and T. Zickler. Statistics of real-world

hyperspectral images. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 193–200. IEEE, 2011. 2

- [5] W. Di, L. Zhang, D. Zhang, and Q. Pan. Studies on hyperspectral face recognition in visible spectrum with feature band selection. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 40(6):1354–1361, 2010. 2
- [6] D. H. Foster, K. Amano, and S. M. Nascimento. Time-lapse ratios of cone excitations in natural scenes. *Vision research*, 120:45–60, 2016. 2

- [7] D. H. Foster, K. Amano, S. M. Nascimento, and M. J. Foster. Frequency of metamerism in natural scenes. *Josa a*, 23(10):2359–2372, 2006. 2
- [8] K. Fotiadou, G. Tsagkatakis, and P. Tsakalides. Deep convolutional neural networks for the classification of snapshot mosaic hyperspectral imagery. 2
- [9] B. Geelen, N. Tack, and A. Lambrechts. A compact snapshot multispectral imager with a monolithically integrated per-pixel filter mosaic. In *Spie Moems-Mems*, pages 89740L–89740L. International Society for Optics and Photonics, 2014. 3
- [10] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 2
- [11] S. T. Namin and L. Petersson. Classification of materials in natural scenes using multi-spectral images. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pages 1393–1398. IEEE, 2012. 2
- [12] M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Ng. Ros: an open-source robot operating system. In *Proc. of the IEEE Intl. Conf. on Robotics and Automation (ICRA) Workshop on Open Source Robotics*, Kobe, Japan, May 2009. 5
- [13] T. Scharwächter, M. Enzweiler, U. Franke, and S. Roth. Efficient multi-cue scene segmentation. In *German Conference on Pattern Recognition*, pages 435–445. Springer, 2013. 2
- [14] T. Skauli and J. E. Farrell. A collection of hyperspectral images for imaging systems research. In *Digital Photography*, page 86600C, 2013. 2
- [15] G. Tsagkatakis, M. Jayapala, B. Geelen, and P. Tsakalides. Non-negative matrix completion for the enhancement of snap-shot mosaic multispectral imagery. 2016. 3
- [16] G. Tsagkatakis and P. Tsakalides. Compressed hyperspectral sensing. In *SPIE/IS&T Electronic Imaging*, pages 940307–940307. International Society for Optics and Photonics, 2015. 3
- [17] G. Tsagkatakis and P. Tsakalides. A self-similar and sparse approach for spectral mosaic snapshot recovery. In *2016 IEEE International Conference on Imaging Systems and Techniques (IST)*, pages 341–345, Oct 2016. 3
- [18] G. Tzagkarakis, W. Charle, and P. Tsakalides. Data compression for snapshot mosaic hyperspectral image sensors. 2016. 3
- [19] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar. Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE transactions on image processing*, 19(9):2241–2253, 2010. 2
- [20] Y. Zhang, C. P. Huynh, N. Habili, and K. N. Ngan. Material segmentation in hyperspectral images with minimal region perimeters. *IEEE International Conference on Image Processing*, September 2016. 2