eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

The HSI-Drive 2.1 dataset

White Paper (2025)

GDED Group

Authors	Project
Koldo Basterretxea	
Jon Gutiérrez Zaballa	HSI-Drive

Bilbao, May 23, 2025

Contents

1 | Introduction

HSI-Drive dataset was created with the aim to contribute to the research of the use of Hyperspectral Imaging (HSI) in the development of intelligent vision for Autonomous Driving Systems (ADS). The seminal idea was that, due to the availability of new small-form factor, portable snapshot HSI cameras, it was worth investigating whether this technology could help develop more efficient and robust vision systems for ADS. More efficient in the sense that, by acquiring richer information about the spectral reflectance properties of the materials in an image, it would be possible to simplify the back-end algorithms responsible for interpreting the world around the vehicle (scene understanding) and thus optimize processor design and reduce computational burden. More robust because the use of spectral information, together with the augmentation of the sensed light spectral bandwidth (VIS/NIR/SWIR), could help to overcome some limitations of RGB imaging vision systems such as metamerism and performance loss under adverse lighting and weather conditions.

However, before using this dataset in the research for new HSI-based computer vision systems, it is of the utmost importance to correctly understand the nature of these data. Firstly, it is necessary to understand the fundamentals of the technology that makes it possible to produce relatively cheap HSI sensors that allow for the commercialization of snapshot cameras that can record images at video rates, and thus suitable for applications such as ADS. Secondly, it must be taken into consideration that the uncertainty in environmental conditions inherent to obtaining HSI video in real driving scenarios (light, vibrations, relative speed of objects in a scene, etc.), especially if the approach is a simple setup that reproduces a realistic implementation situation, has a profound impact on the quality of the data obtained. The combination of these factors makes the image data provided in this dataset far from what would be obtained with a high-end spectroscopic camera under laboratory conditions, and therefore should not be interpreted or used as such. The objective of this document is to provide relevant, while not exhaustive, information on the characteristics of the HSI-Drive data and thus help researchers who decide to use this dataset to interpret them correctly and to apply them in accordance with their nature.

The v2.1 version of the HSI-Drive dataset does not provide more annotated images, but a new, more careful annotation of the images already in version v2.0. The aim of this new labeling effort has been twofold. Firstly, to increase the amount of labeled pixels for training, especially in the most underrepresented categories. Secondly, to provide higher quality test images for the evaluation of segmentation models. However, the primary approach to the image labeling of the dataset is not changed, i.e. keeping unlabeled the pixels that a human labeler cannot clearly decide which category they belong to. This usually includes many background pixels and all the edges that delimit different items or surfaces in a scene.

Table 1.1: Frequency of each class in the HSI Drive v2.0 dataset.

	Total	Road	R.Marks [a]	Veg.[b]	Pain.Met. [c]	Sky	Concrete	Ped. [d]	Water	Unpain.Met. [e]	Glass
Pixels	43,947,503	26,690,619	1,325,343	9,339,224	948,852	2,511,496	2,315,153	209,531	12,330	348,341	246,614
%	100	60.73	3.02	21.25	2.16	5.71	5.27	0.48	0.03	0.79	0.56

[a]Road Marks. [b] Vegetation. [c] Painted Metal. [d] Pedestrian. [e] Unpainted Metal.

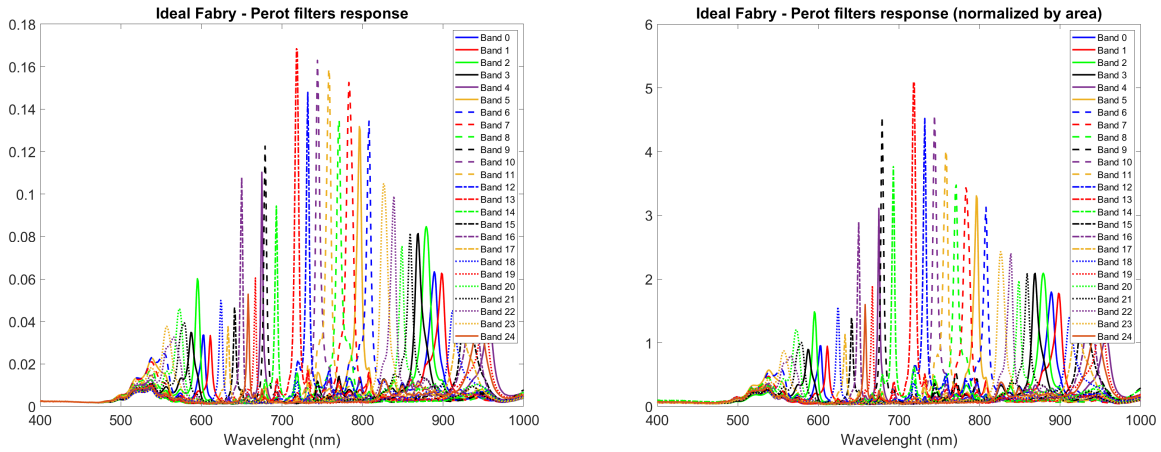
Table 1.2: Frequency of each class in the HSI Drive v2.1 dataset.

	Total	Road	R.Marks [a]	Veg.[b]	Pain.Met. [c]	Sky	Concrete	Ped. [d]	Water	Unpain.Met. [e]	Glass
Pixels	45,055,512	26,753,811	1,364,908	9,799,475	1,113,573	2,549,527	2,485,658	231,019	10,592	467,688	279,261
%	100	59.38	3.03	21.75	2.47	5.66	5.52	0.51	0.02	1.04	0.62

[a]Road Marks. [b] Vegetation. [c] Painted Metal. [d] Pedestrian. [e] Unpainted Metal.

2 | HSI sensor technology

The images contained in the HSI-Drive dataset were recorded with a Photonfocus MV1-D2048x1088-HS02-96-G2 camera that features an Imec 25-band HSI NIR sensor. The spectral bands are obtained by a mosaic of Fabry-Perot (FP) filters deposited on a CMOSIS CMV2000 image wafer sensor that produces 2D images with 5x5 pixel windows. According to Imec's technical documentation, the active range of the sensor is 600-975 nm, i.e. it is a red-NIR sensor. However, it is important to note that to limit the sensor response to this wavelength band, rejection filters have to be mounted in front of it. If not, the sensor response without rejection filters is that shown in Fig. 2.1. Notice that there is a spectral leakage below the 600nm limit and that many filters present a second-order response peak. In the case of the Photonfocus MV1 camera, it does not feature an internal high-pass 600nm filter integrated with the sensor, which implies using band-pass rejection filters attached to the optics in case low-wavelength leaking and mixed first or second order responses want to be filtered out [1]. However, as shown in Table 2.1, this is not possible to achieve for all the 25 FP filters in the mosaic.



(a) Ideal Fabry-Perot response

(b) Ideal Fabry-Perot response (normalized by area)

Figure 2.1

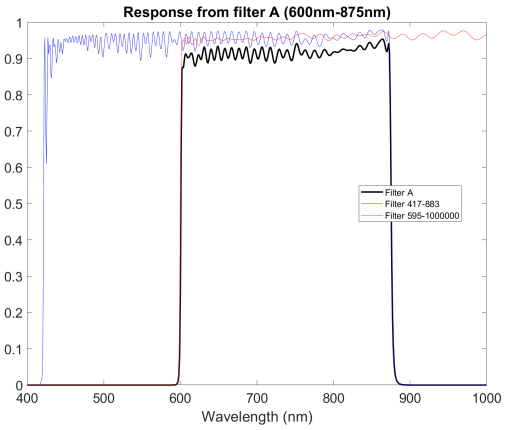
Table 2.1 shows the contribution of the peak signals to the total band response when no rejection filters are used. It is measured as the area under the fitted peak in [center ± 1.5 x FWHM] relative to the total response of the sensor.

Fig. 2.2 shows the frequency responses of the two rejection filters provided with the Photonfocus camera. We named filter A the rejection filter that provides a [600nm, 875nm] pass band, and filter B the filter that provides a [675nm, 975nm] pass band.

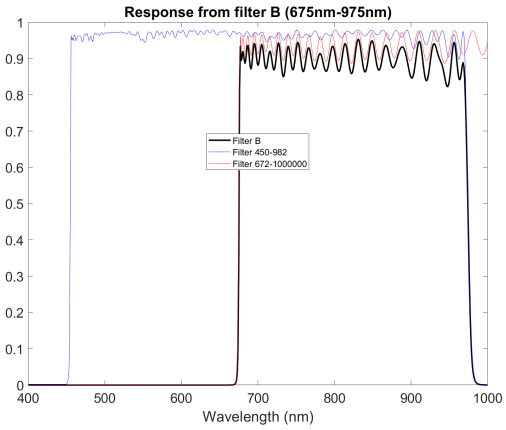
[1] In latest versions of the Photonfocus cameras a bandpass filter is integrated inside of the camera body, eliminating the need to mount rejection filters on the lens

	First order				Second order			
Filter	Wavelength (nm)	Contribution	FWHM (nm)	Quant. Eff. (%)	Wavelength (nm)	Contribution	FWHM (nm)	Quant. Eff. (%)
0	888.479	0.289	12.91	6.3	601.711	0.094	8.08	3.3
1	897.659	0.270	12.54	6.1	610.953	0.072	6.40	3.2
2	879.132	0.334	12.54	8.6	595.124	0.124	6.40	6.2
3	869.114	0.279	10.87	8	587.481	0.127	11.80	3.3
4	956.111	0.230	16.63	3.9	675.021	0.124	3.06	11.5
5	795.962	0.284	6.77	13.4	539.053	0.177	25.00	2.3
6	807.631	0.259	6.77	13.1	552.310	0.145	19.42	2.6
7	783.757	0.412	9.19	15.9	537.869	0.137	23.51	2.1
8	770.576	0.320	7.15	13.9	536.812	0.123	21.84	1.7
9	679.136	0.239	4.36	11.8	770.561	0.059	9.19	1.4
10	744.483	0.331	6.03	15.6	535.563	0.101	23.88	1.2
11	757.923	0.346	6.78	16.1	536.038	0.108	23.51	1.4
12	731.883	0.246	4.36	14.7	783.642	0.064	12.91	1.3
13	718.298	0.346	5.10	17.9	770.861	0.045	10.50	1.1
14	692.903	0.181	3.99	9	744.417	0.044	7.71	1.1
15	928.810	0.368	21.47	4.5	641.358	0.075	4.36	4.6
16	936.246	0.357	19.05	5.6	649.777	0.126	3.62	10.4
17	920.599	0.292	14.96	5.2	632.874	0.065	4.73	3.7
18	912.169	0.243	14.21	4.4	624.181	0.095	4.73	5.1
19	950.391	0.289	16.63	4.4	666.719	0.076	3.06	6.3
20	848.549	0.192	8.08	7.3	572.240	0.177	11.24	4.8
21	858.948	0.222	8.82	7.9	577.147	0.175	13.29	4.1
22	564.362	0.246	25.00	3.2	838.080	0.238	8.08	9.7
23	826.768	0.329	10.50	10.8	557.051	0.179	16.26	3.8
24	944.485	0.347	22.77	4	657.995	0.068	3.24	5.5

Table 2.1: Filter response contribution for each of the 25 FP filters in the mosaic



(a) Filter A response



(b) Filter B response

Figure 2.2: Frequency responses of the rejection filters

As an example, in Fig. 2.3 we show the graphical representation of the first FP filter when, top, Filter A is used, center, no filter is used, and bottom, Filter B is used.

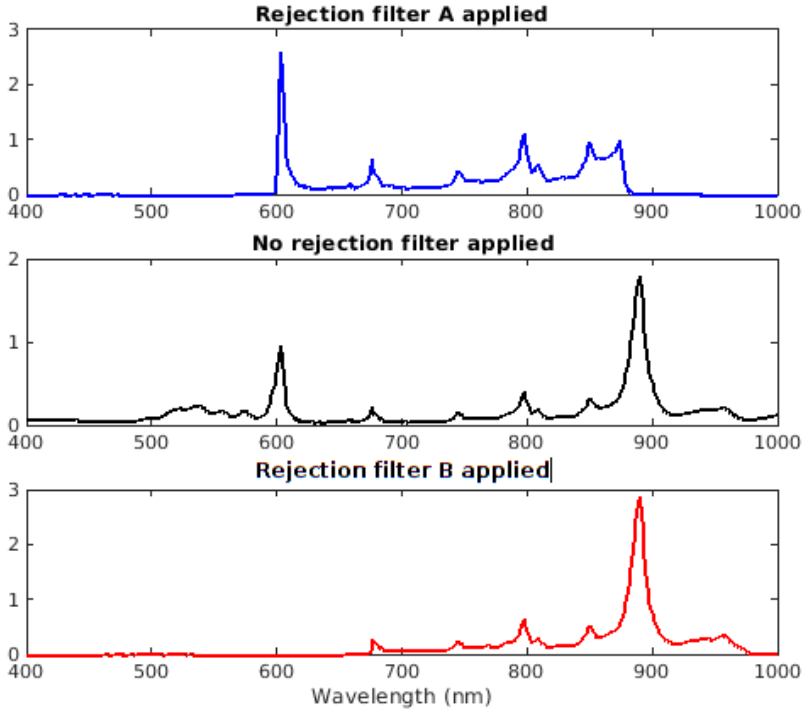


Figure 2.3: Normalized response of the first Fabry-Perrot filter (band 0): peaks at 880.5 nm and 601.7 nm

As can be observed, rejection filters improve the separability of the spectral information, avoiding spectral leakage under the 600nm limit and minimizing second-order peaks. However, even with the use of rejection filters, there is still an unavoidable spectral mixing in the sensor response. This spectral mixing, which is present to a greater or lesser extent in every FP filter, can be reduced in the postprocessing of the acquired raw image by means of a linear transformation matrix (spectral correction), as far as rejection filters are used in the image acquisition.

3 | Recording setup

The recordings for the HSI-Drive dataset were performed with no rejection filters attached to the optics of the camera. This decision, which may be controversial, was motivated by two main factors.

Figure 3.1: The Photonfocus snapshot HSI NIR camera mounted on a car

Firstly, because rejection filters notably reduce the amount of light reaching the sensor and also produce severe vignetting. This fact is of great importance for this application due to the demanding requirements in terms of image integration-time and depth of field. On the one hand, image integration time was initially limited to 10,000us to avoid motion blur. This limit had to be later modified, at the risk of producing some blurring for fastest moving elements in the scenes, to 20,000us for recordings under low-illumination conditions. On the other hand, the calculated optimum focal aperture to get the desired depth of field was f=8.0. Again, this constraint had to be later relaxed to f=4.0 for low-illumination conditions. The combination of these two factors define de Exposure Value (EV) of each frame shot. The use of rejection filters would have required the use of even larger apertures, unacceptably reducing the depth of field and degrading the functionality of the FP filters.

Secondly, because the initial exploring experiments performed with Machine Learning (ML) baseline spectral classifiers (ANN and SVNs) on images recorded with and without rejection filters did not show neither notable nor a consistent improvement in classification performance for the former setup. All in all, the final decission was to priotitize the practical application approach over the quality of data from a spectral analysis perspective and avoid the use of rejection filters [2].

Finally, it has to be mentioned that in some recording sessions the Analog Gain parameter of the Photonfocus camera was set to x2.0 to further amplify the sensor response (at the cost of augmenting the noise). The combination of these three parameters, integration time (t), focal aperture (f) and analog gain (AG), resulted in eight different recording setups or camera configurations, which has implications in the postprocessing of the raw images necessary to get the 3D spectral cubes. Table 3.1 shows the details of the eight applied camera setups and the calculated relative equivalent exposure (EV) and the absolute equivalent exposure (AEV) numbers for each configuration. As can be seen, there are five different AEV

[2]While with the NIR sensor the out of band leakage of the filters is limited and thus the spectral signal is not too much contaminated, the choice to not use the rejection filters makes an accurate irradiance or reflectance spectrum can no longer be computed from the data. While this is not a requirement for the classification, in case accurate spectral measurements are required the bandpass filters should not be removed

values from 8.64 to 12.64, which implies a iluminance variability factor of x16 between configurations 1 and 6.

Table 3.1: Camera setups

setup	time us	f	EV	AG	AEV
1	10000	8	12.64	1	12.64
2	10000	8	12.64	2	11.64
3	10000	4	10.64	1	10.64
4	10000	4	10.64	2	09.64
5	20000	8	11.64	1	11.64
6	20000	8	11.64	2	10.64
7	20000	4	09.64	1	09.64
8	20000	4	09.64	2	08.64

4 | Cube processing pipeline

As explained in Section 2, the sensor produces 2D images with a spectral mosaic pattern. If 3D spectral cubes are to be generated from raw images, some image processing steps must be performed. Explaining the very details of the cube generation processing pipeline is out of the scope of this document, and anyway, there is no standardized cube processing pipeline; it depends on the goals, the processing constraints, and the application domain. With this in mind, the HSI-Drive dataset includes the original binary RAW images (12 bit resolution) together with the generated 3D spectral cubes.

Following the same experimental philosophy that guided the recording setup, i.e. prioritizing the practical application approach, the cube processing pipeline applied to generate the hyperspectral cubes in the HSI-Drive 2.1 was kept simple. Firstly, because the used recording setup (outdoor recording, different camera setups, etc.) strongly conditions the applicable postprocessing approach. Secondly, because the target application (ADS) requires real-time processing of the acquired images; trying a more accurate yet more complex cube generation pipeline would become the principal processing bottleneck that would impede it.

The applied spectral cube generation process is a reflectance processing pipeline that comprehends the following steps (see Appendix A):

1. Image cropping and framing: It is necessary since the mosaic filter does not cover the whole sensor.
2. Bias and reflectance correction: A dark field image and a white field image are used to compensate for sensor noise and perform white balancing. Ideally, this process should transform the irradiance information into a more or less accurate spectral reflectance information. However, this cannot be achieved with the recording conditions used to record the images in this dataset. The v2.1 version incorporates a new processing function (optional) to estimate the scene relative illumination with respect to the reference white field images and perform a scaling of the white balancing. Further explanations on this issue are given below.
3. Partial demosaicing: It is the process of selecting the pixels that correspond to the same spectral band at each mosaic filter. This implies reducing the image resolution to 1/5.
4. Spatial filtering (optional): For coherence with the previous versions of the dataset, we provide cubes generated with and without 3x3 median spatial filtering (only for the cubes with no white balancing).
5. Translation to center (band alignment): A bilinear interpolation algorithm is applied to compensate for the spatial offset of the pixels surrounding the central pixel in the mosaic filter.
6. Data normalization (optional): Since data normalization techniques (per band normalization, pixel normalization etc.) have different objectives depending on the algorithms to be used to subsequently process the hyperspectral images, and since it is the final step of the processing pipeline, unlike in the v1.x versions of the dataset, we do not provide normalized cubes in v2.x versions. This process, if necessary, is left to the dataset users.

Notice that no spectral correction stage is applied. As detailed in Section 3, no rejection filter was used for recording, thus no spectral correction matrix is available to compensate for sensor spectral mixing.

Reflectance Correction: Reflectance correction is aimed at canceling the irradiance spectrum of the illuminant. In a laboratory, this can be performed by using a calibrated reference white tile to divide the data in the image to be processed by the data in a reference white image acquired under the same illumination conditions and with the same camera configuration. In outdoor scenes this is not possible, but it may be sufficient that the reference tile is visible in part of one of the scenes assuming that illumination is stable. If not, as is the case for the recording of the HSI-Drive dataset scenes, an accurate reflectance correction is not possible at all. However, when generating data for the HSI-Drive dataset, we chose to use a common white reference image obtained for each of the four different f-number/AG combinations to perform a ”pseudo reflectance correction”. These white reference images were generated by averaging various shots over a spectrally calibrated Spectralon white reflectance tile under expected natural maximum illumination conditions, i.e at midday on a sunny day (Fig. 4.1. Since no point spectrometer and no other

light measuring devices were present in the recording setup to compensate for illumination variations, obviously the applied correction does not normalize the spectral signatures of the different images in the dataset (a posteriori image normalization techniques can be applied to deal with this to some extent). However, this processing still provides some benefits. Firstly, it reduces sensor non-uniformity issues and image vignetting, and secondly, it cancels to some extent the irradiance spectrum of natural light.

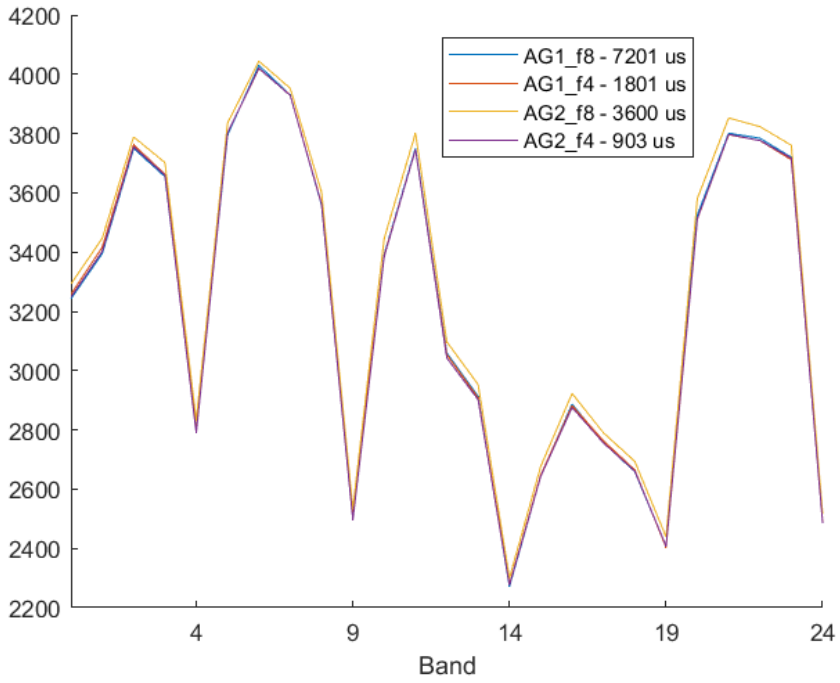
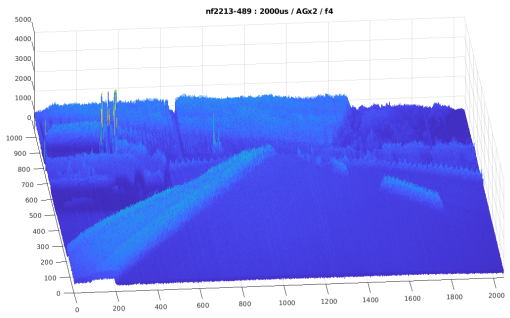
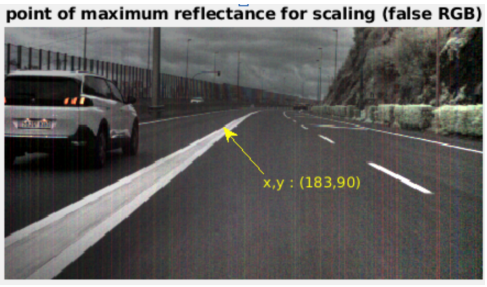


Figure 4.1: Spectra of the maximum values of the averaged reference white tile images for the four different camera configurations used in the dataset. Images were taken by exposing the calibrated tile to direct sunlight in a clear day with the sun at its zenith

As an improvement to data quality, in the v2.1 version of the dataset we incorporate an additional processing function that estimates the relative level of illumination of the recorded scene by searching for the pixels with the highest albedo at each image. These pixels usually correspond to high-reflectance white surfaces such as road marks, white vehicle bodies, etc. although in some cases the algorithm selects pixels corresponding to the sky. By comparing the irradiance of these pixels with the reference white images, a scaling factor is calculated to correct the reference white images stored in memory. Ideally, if the procedure was perfect, all images in the dataset would be scaled in the [0,1] range. However, the search for reference pixels for scaling involves rejecting pixels from artificial light sources such as illuminated signs, traffic lights, and front and rear lights of vehicles. The programmed algorithm automatically segregates these ”suspicious” pixels on the foundation of their spectral signatures, so no human intervention is required and thus, this function can be embedded in the image processing pipeline of the image segmentation processor (see Fig. 4.2). Artificial light pixels are thus treated as outliers and clipped to 1 at the end of the cube preprocessing sequence.



(a) Sensed irradiance values (int12)



(b) False RGB image of the 25-band HSI cube and coordinates of the pixel with maximum albedo

Figure 4.2: Example of the automatic identification of a maximum albedo pixel for the white balance scaling. This image corresponds to a cloudy morning fall recording with low lightning. Although the maximum irradiance is generated by the rear and front lights of the cars (a), the algorithm successfully rejects these pixels and selects a pixel corresponding to the road mark as the highest refelctance pixel in the image for scaling (b)

Performed testing experiments on the HSI-Drive 2.1 dataset demonstrates the superiority of applying this scaled reflectance correction processing function over the use of non-scaled cubes both with and without applying pixel normalization procedures. Table 4.1 summarizes the results obtained in the HSI-Drive 2.1 dataset for a 5-class experiment with our latest and best performing model, which is a U-Net with spectral attention modules. It can be seen that the best global results are obtained for the scaled reflectance correction version with no pixel normalization (PN). Notice that when no scaling strategy was applied to the reflectance correction stage, as reported in previous publication, applying PN produced better results compared to the non normalized data. Nevertheless, in the v2.1 version of the dataset both scaled and non-scaled cubes are provided for experimentation.

Table 4.1: Segmentation results for the 5 class experiment. The figures correspond to mean IoU values over a 5-fold cross-validation experimental setup

version	road	road marks	vegetation	sky	”others”	global	weighted
No scaling + PN	97.64	85.33	94.55	92.89	81.79	94.71	87.75
Scaling + PN	97.83	87.27	94.60	94.14	82.50	95.04	89.16
Scaling	98.04	89.97	94.46	92.05	83.26	95.17	90.03

Figure 4.3: Example of a segmented image for the 5-class experiment. Cube processing: scaled reflectance correction and no pixel normalization.

A | Appendix A: Processing pipeline

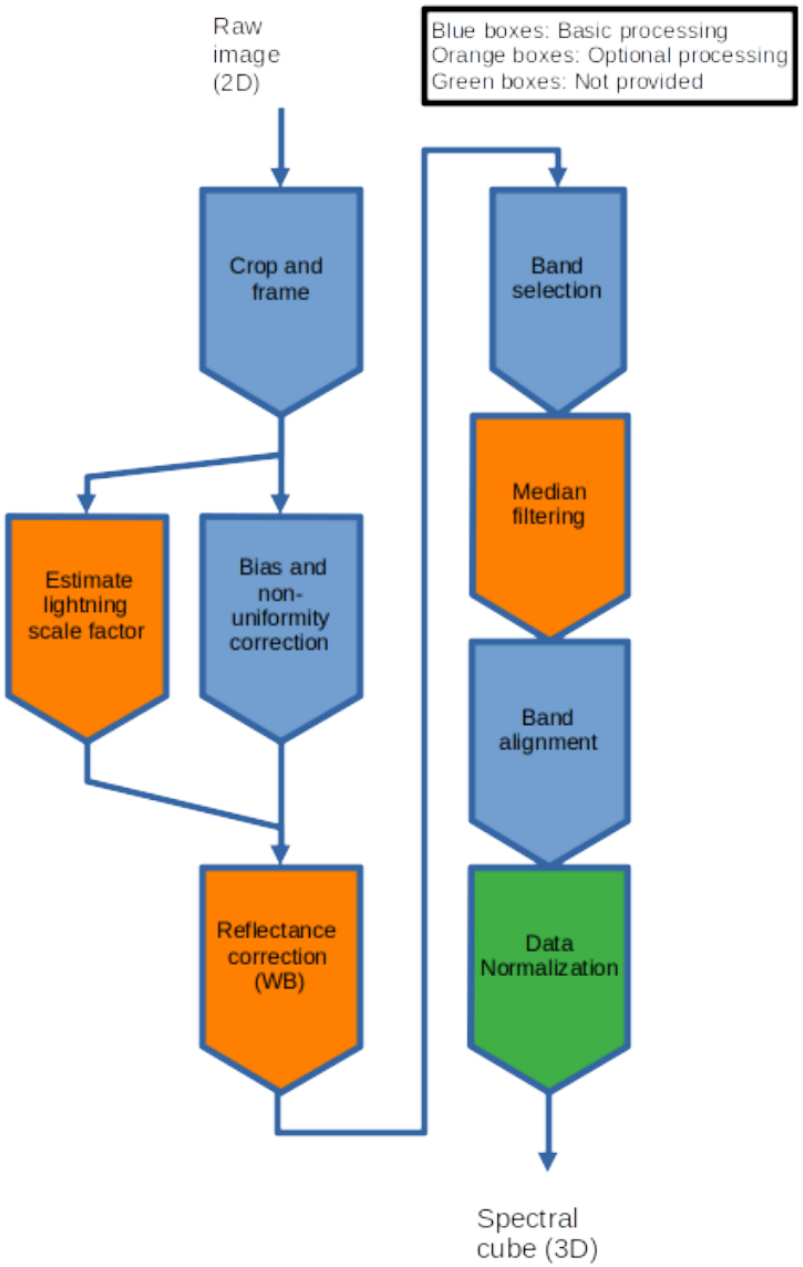


Figure A.1: Pipeline flow for the HSI-Drive v2.1 dataset cube processing