Autonomous natural landmarks identification for outdoor navigation

Caterina Massidda*, Heinrich H. Bülthoff*† and Paolo Stegagno*

Abstract— **Identification of landmarks for outdoor navigation is often performed using computationally expensive computer vision methods or via heavy and expensive multi-spectral and range sensors. Both choices are forbidden on Micro Aerial Vehicles (MAV) due to limited payload and computational power. However, an appropriate choice of the hardware sensor equipment allows the employment of mixed multi-spectral analysis and computer vision techniques to identify natural landmarks. In this work, we propose a low-cost low-weight camera array with appropriate optical filters to be exploited both as stereo camera and multi-spectral sensor. Through stereo vision and the Normalized Difference Vegetation Index (NDVI), we are able to classify the observed materials in the scene among several different classes, identify vegetation and water bodies and provide measurements of their relative bearing and distance from the robot. A handheld prototype of this camera array is tested in outdoor environment.**

I. INTRODUCTION

Recent advances in robotics have made possible a growing number of real world applications of robotics systems in structured environments, in which external sensors (e.g.: motion capture devices) or artificial *a priori* known landmarks are available. However, in order to employ robotic systems in unstructured outdoor environments, the robots should be able to autonomously identify and select natural landmarks as for example trees and water bodies and collect some relative measurement as bearing and distance. Among other tasks, this knowledge can be used to achieve localization, plan the action or perform visual servoing.

In most previous works, the identification of particular features in the environment as roads and trees is performed using computationally expensive computer vision algorithms, either employing only cameras or in combination with some range sensor. In one of the first papers on this topic [1] the authors fuse the information gathered from a sonar and a camera to measure the relative position of a tree. However, their method requires many assumptions on the shape of the trees and is suitable only in particular situations. In [2] the authors employ color segmentation in order to extract roads from an image. In [3], trees are identified using edge extraction and circular features identification. Many works [4], [5], [6] focus on the identification of most salient regions of an image in order to select landmarks for autonomous navigation, independently from the actual type of identified landmark. In [7], a classification approach has been implemented in order to extract tree trunks from

the camera images. The authors of [8] have employed a contrast-based and shape estimation method to estimate the position and the size of trees in the scene through the images gathered by an omnidirectional camera. With the advent of 3D scanners and RGB-D sensors, similar algorithms (shape identification, clustering) have been applied [9] on 3D point clouds.

The problem of the identification of vegetation, water bodies and other materials from satellite and airborn images is well known in remote sensing. However, it is usually solved exploiting a completely different approach based on the measurement of the reflectance of the Electro-Magnetic (EM) radiation at different wavelengths. This is done collecting multiple images of the same scene at different wavelengths, not only in the visible spectrum but also in the ultra-violet and in the infrared. In [10], the authors have employed this principle to identify the vegetation for an autonomous navigation system. In particular, they have used a cold mirror to split the visible and the Near Infrared (NIR) part of the light among two cameras, and the two resulting images are used to compute the pixel-by-pixel Normalized Difference Vegetation Index (NDVI, see Sect. II). However, their setup, described in [11], includes many additional sensors as sonars and laser scanners and is not suitable for the application on MAVs.

In this paper, we are interested in this second approach. Being our reference platform an MAV, we aim at integrating a low-cost low-weight multispectral sensor in a landmark detection system. Unfortunately, on-the-shelf multi-spectral sensors are still quite expensive and heavy, hence not suitable for our purpose. In addition, we would also like to obtain metric information (e.g.: distance, size) of the observed objects, but a normal spectral sensor does not provide it. In a normal setup, as the one proposed in [10], metric information can only be retrieved by means of additional range sensors, but the limited payload of MAVs does not allow this option. For this reason, we have decided not to rely on the principle of the cold mirror to design our sensor.

Another interesting attempt to build a low-cost spectral camera relies on the idea of creating an array of cameras and doting each of them with a different optical filter [12]. Similarly, the authors of [13] outfitted a fixed-wing UAV with an array of sensors including a multi-spectral camera. One issue that must be addressed in this type of sensor arrays is the matching between corresponding points of the images. In fact, in order to extrapolate useful data, it is necessary to identify all the pixels in the different images that refer to the same physical object. Unfortunately, this matching is highly dependent from the distance of the observed object. Hence,

*C. Massidda, H. H. Bülthoff and P. Stegagno are with the Max Planck Institute for biological Cybernetics, Tübingen, Germany. E-mail: {caterina.massidda, hhb, paolo.stegagno}@tuebingen.mpg.de.

†H. H. Bülthoff is also with the Department of Brain and Cognitive Engineering, Korea University, Seoul, 136-713 Korea.

without the knowledge of this distance it is not possible to correctly match the pixels in the images collected at different wavelengths.

In this work, we propose a sensor design to use the same sensor both as stereo camera and multi-spectral camera, so that we can simultaneously solve the problem of the pixel association in the spectral camera and obtain metric information. We use two cameras as a stereo pair and compute the corresponding disparity and depth maps. Using this information, we are able to reconstruct the point of view of the other camera(s), and compute the correct matching for each pixel of the image. The matched images will then be used to classify the observed objects in a given number of different classes, and finally identify landmarks as objects in the environment of a certain type. The final outcome is a point-cloud of the landmarks that can be then used in the different tasks of outdoor navigation. For the best of our knowledge, our system is the first to integrate multi-spectral classification and stereo vision in a single low-cost low-weight sensor that is suitable for the application on MAVs.

The rest of the paper is organized as follows, Section II introduces some background on multi-spectral analysis and the NDVI. Section III describes our hardware setup, while Section IV describes the software used to make the online unsupervised classification. Sections V and VI present respectively the experimental results and the conclusions where we describe the future applications and improvements.

II. BACKGROUND ON SPECTRAL ANALYSIS

Multi-spectral analysis is a powerful technique to classify materials currently applied in many fields as topography, geology, archeology, meteorology among others. It is based on the physical principle that the percentage of electromagnetic (EM) radiation reflected by a given material (reflectance) varies with respect to the wavelength of the incident radiation. The reflectance of a material constitutes a sort of 'unique spectral signature', and can be used to identify the material itself by comparing online collected data with databases of known materials.

Some significant examples of spectral signatures are reported in Fig. 1. The knowledge of such spectral signatures is used to identify the type of observed surfaces. Ideally, measuring the whole spectral signature of a surface and matching it with the spectral signatures of known materials will allow its identification. However, sensors able to measure enough points to build the whole spectral signature of an object, known as hyper-spectral cameras, are even more expensive and heavy than multi-spectral cameras. On MAVs, the sole option is to rely only on few wavelengths

When working under this condition, one technique of spectral analysis consists in developing some mathematical combination of the collected values to accentuate the spectral properties of the targeted materials. Among those indexes, the NDVI was developed by NASA to identify vegetation and study its status of health. Let be RED and NIR respectively the spectral values relative to a surface in the red and Near Infra Red (NIR) bands, then the general equation for the
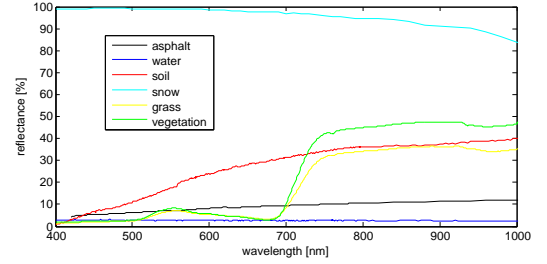

Fig. 1: Spectral signatures of some materials [15].

NDVI is:

$$\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}} \qquad (1)$$

in which the difference of the NIR and RED values is normalized by their sum. The normalization scales the NDVI to the set $[-1, 1]$ and reduces the influence of atmospheric absorption and the artifacts related to sensor noise.

Since the NDVI is largely used, it is possible to retrieve in literature many values of common materials of outdoor environment, e.g. [14]. A classification largely accepted in the scientific world differentiate the materials in four main macro-areas ω_c, $c = 1, \ldots, 4$, for which some examples of spectral signatures are reported in Fig. 1.

The first macro-area ω_1 includes a wide variety of healthy vegetation, and account for the strong absorption in the red band due to clorophilla and the equally strong reflection of the NIR EM radiation of the green vegetation. The two effects result in high values of the NDVI index ($\text{NDVI}(\omega_1) > 0.6$). Low and dry vegetation fall in the category ω_2, in which the two phenomena are less strong, hence the NDVI values ($0.2 < \text{NDVI}(\omega_2) < 0.6$) are still positive but less than the previous category. Bare soil, as well as concrete and some other man-made materials, have in general a slightly increasing spectral signature (Fig. 1), while the spectral signature of asphalt is nearly flat. The corresponding NDVI values ($0 < \text{NDVI}(\omega_3) < 0.2$) are still positive but close to zero. Finally, water in lakes, rivers and seas, as well as snow, have a stronger absorption in the NIR band with respect to the red band, resulting in negative NDVI values ($\text{NDVI}(\omega_4) < 0$). This classification is summarized in the following equation:

$$
\begin{aligned}
&\text{NDVI}(\omega_1) > 0.6 \\
&0.2 < \text{NDVI}(\omega_2) < 0.6 \\
&0 < \text{NDVI}(\omega_3) < 0.2 \\
&\text{NDVI}(\omega_4) < 0.
\end{aligned}
\qquad (2)
$$

III. HARDWARE SETUP

The camera array consists of three USB 2.0 mvBlueFOX grayscale cameras mounted in a straight line on a 3D printed frame shown in Fig. 2a. The cameras have a resolution of 752x480 pixels and offer a complete control to the user, allowing for example external triggering and the setup of the exposition time. In the following, we will refer the left,

central and right cameras respectively as $CamL$, $CamC$, and $CamR$. $CamC$ is positioned in the middle between $CamL$ and $CamR$.

Each camera is equipped with an S-mount low distortion ($dis < 0.2\%$) Lensagon B5M41430ND lens. The $CamL$ and $CamR$ lenses are equipped with a dark red bandpass filter (wavelength = 660 [nm]) with a diameter of $d = 22.5[mm]$. The $CamC$ lens is equipped with an infrared bandpass filter (wavelength = 850 [nm]) with the same diameter. The frame with the cameras, lens and filters are shown in Fig. 2b. This layout allows us to use the $CamL$ and the $CamR$ as stereo camera.

An important factor in the choice of the camera and the lenses is their efficiency with respect to the different wavelengths. In particular, we considered the quantum efficiency QE of the camera sensor and the transmission efficiency of the filters. The two contributions must be considered together in order to obtain the total efficiency of the filter/camera sensor system. The QE is related to the ability of the sensor to respond to the incoming photon signal and to its conversion into a measurable electron signal. The QE is usually expressed as the probability that a photoelectron will be released for each incident photon, and is in general a function of the wavelength λ of the incident light ($QE \triangleq QE(\lambda)$). The chosen sensor is sensible in the bands of interest (red and NIR).

The transmission efficiency of the filters is the ratio of the transmitted power over the incident power, and is again a function of the wavelength λ. In our case, we refer to the transmission efficiency of the red and NIR filters respectively as TE_{Red} and TE_{NIR}. The plots of the total efficiency of the red and NIR filters/camera systems with respect to the wavelength λ, respectively $E_{Red}(\lambda) = QE \cdot TE_{Red}$ and $E_{NIR}(\lambda) = QE \cdot TE_{NIR}$, are shown in Fig. 3, along with QE, TE_{Red} and TE_{NIR}.

The three cameras are connected through USB ports to an i5 PC which runs the software to acquire the camera images and performs the unsupervised classification of the observed objects. The overall weight of the 3D printed frame, cameras, lenses and filters is $\simeq 100g$, while the total cost of those components is below 1000 Euros.

IV. SOFTWARE SETUP

The software to acquire the camera images and perform the unsupervised classification encompasses several components. A block scheme of the whole system is provided in Fig. 4, and will be the reference for the reader throughout this Section. In general, we will denote the images gathered from the left, central and right cameras respectively as $RedL$, $NIRC$ and $RedR$. Each camera captures a frame in the same instant, collecting three images from three different points of view of the same scene.

A precondition to perform the classification is to ensure that the images contains useful information by an appropriate selection of the exposure times of the cameras. If the exposure of a camera sensor is too long the gathered image will be all white, while a too short exposure will
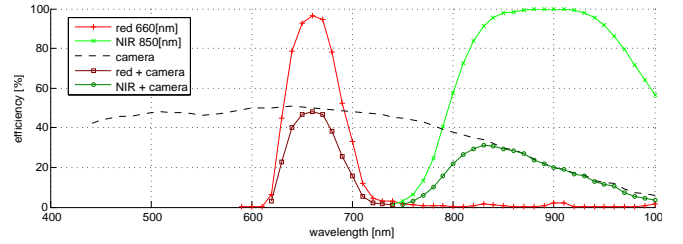

Fig. 3: Widths and efficiency of the two filters used in the multispectral sensor. The light red and dark red plots show respectively the transmission efficiency of the red filter and of the red filter mounted on the camera. The light green and dark green plots show respectively the transmission efficiency of the NIR filter and of the NIR filter mounted on the camera. The dark dashed line shows the quantum efficiency of the camera sensor.

result in a dark image. Since $CamR$ and $CamL$ will be used as a stereo pair, it is in general good practice to select the same exposure time ExT_{Red} for both cameras, while the exposure time ExT_{NIR} of $CamC$ can be set independently. Thereafter, we implemented two independent PID controllers to regulate the brightness of the images, using as input the mean value of the pixels of the images.

Once the images are acquired, they are passed on to the system responsible for the classification. The first step is to find the correspondences among the pixels of $RedL$ and $RedR$ and the pixels of $NIRC$, or equivalently to reconstruct with $RedL$ and $RedR$ the same point of view of $NIRC$.

As mentioned in Section III, this is done by using $CamL$ and $CamR$ as a stereo camera. The software computes the disparity map D_{Red} of $RedL$ and $RedR$ and uses it to reconstruct the same point of view of the central camera $CamC$. In this way, we have two images $RedC(RedL, RedR)$ and $NIRC$ taken from the same point of view with two different bands. From $RedC$ and $NIRC$ it is possible to compute the pixel by pixel NDVI value to obtain the NDVI matrix, denoted by $NDVI$, as we will show in Section IV-A. This can be used to classify the surfaces according to standard values known from literature.

However, since $NDVI$ is usually corrupted by consistent noise (due to e.g.: mismatches in the association, difference in the illumination, etc.), the classification is conducted considering spatial information in addition to spectral information. The application of such clusterization algorithm, which we will explain in detail in Section IV-B and is based on [16], leads to a consistent reduction of outliers and erroneous classifications.

A. NDVI

The matrix $NDVI$ cannot be computed directly on the pixel values of $RedC$ and $NIRC$. In fact, two more factors must be considered to apply equation (1). First, since the exposure times are set independently as ExT_{red} for $CamR$
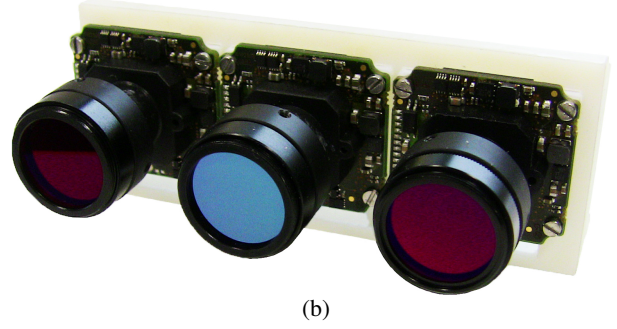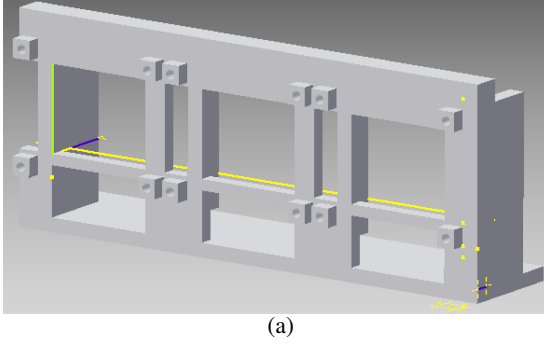
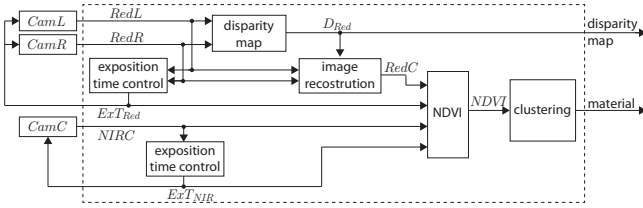Fig. 2: (a) the cad model of the camera support; (b) the camera array with lenses and filters.

Fig. 4: Block diagram of the developed system.

and $CamL$, and as ExT_{NIR} for $CamC$ camera, we need to consider their difference by applying to the values in $RedC$ a multiplicative correction factor δ:

$$\delta = \frac{ExT_{NIR}}{ExT_{red}} \ . \tag{3}$$

The second factor is the difference in the efficiency of the filter/camera systems in the two bands. In particular, the total efficiencies Tot_{red} and Tot_{NIR} of the red filter/camera and NIR filter/camera are respectively:

$$\begin{aligned} Tot_{red} &= \int E_{red} d\lambda = 303.07 \\ Tot_{NIR} &= \int E_{NIR} d\lambda = 433.62 \ . \end{aligned} \tag{4}$$

However, equation (1) requires pure spectral values collected by sensors with the same efficiency in order to be applied. Hence, another multiplicative correction factor α must be applied to $RedC$:

$$\alpha = \frac{Tot_{red}}{Tot_{NIR}} = 1.43 \ . \tag{5}$$

Considering the value of a pixel from the reconstructed image $RedC_{i,j}$ and the corresponding pixel in the $NIRC_{i,j}$ and considering the two correction factors (3) and (5), the pixel-by-pixel NDVI values must be computed as:

$$NDVI_{i,j} = \frac{NIRC_{i,j} - RedC_{i,j} \cdot \alpha \cdot \delta}{NIRC_{i,j} + RedC_{i,j} \cdot \alpha \cdot \delta} \tag{6}$$

where $NDVI_{i,j}$ is the element at the i-th row and j-th column of the matrix $NDVI$.

B. Clustering

Using $NDVI$ computed through equation (6) it is possible to classify the observed materials, i.e.: to assign to each pixel $x_{i,j} \in NDVI$ a label $\omega_{i,j}$ among the set of possible labels

$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ following the rules in equation (2). For visualization purpose, the $\omega_{i,j}$ can be collected in an image Cl whose pixels $Cl_{i,j}$ are defined as a function of $\omega_{i,j}$: $Cl_{i,j} = Cl_{i,j}(\omega_{i,j})$.

However, due to possible mismatches between $RedC$ and $NIRC$, difference in the illuminations, etc., $NDVI$ will be affected by consistent noise and Cl will contain many errors in the classification. Hence, we apply a more robust classification based on the fuzzy K-means clusterization, whose first step consists in computing for each pixel $x_{i,j} \in NDVI$ the membership values

$$P_{spec}(x_{i,j}|\omega_c) = \frac{\Lambda_c(x_{i,j})}{\sum_{h=1}^{4} \Lambda_h(x_{i,j})}, \ c = 1, \dots, 4 \tag{7}$$

where $\Lambda_c(x_{i,j})$ is the probability that a surface of class ω_c produces an NDVI value equal to $x_{i,j}$. Note that $\sum_{c=1}^{n} P_{spec}(x_{i,j}|\omega_c) = 1$, hence $P_{spec}(x_{i,j}|\omega_c), c = 1, \dots, 4$ is a proper probability distribution describing the probability that $x_{i,j}$ was originated by a surface of class ω_c.

In general, the functions $\Lambda_c(x_{i,j})$ can be experimentally estimated. However, given the popularity of the NDVI, it is also possible to retrieve plausible values from literature. The plots of the functions Λ_c that we have used in this work are shown in Fig. 5. In order to analytically describe Λ_c, we have chosen parabolic functions identified by their vertex $(m_c, 1)$ and one point $(b_k, 0.5)$, where $m_c, c = 1, \dots, 4$ are the center of the classes and $b_k, k = 1, 2, 3$ are the transition values from one class to another. The values of m_c and b_k used in this work, retrieved from literature ([14], [15]), are available in Tables Ia and Ib respectively. Other possible choices for the analytical description of the functions Λ_c are for example sigmoid functions. However, the computation of sigmoid functions requires more time w.r.t. quadratic functions and should be avoided.

The row $P_{spec}(x_{i,j}|\omega_c)$ probabilities are used as initialization values for the membership probability $P(\omega_c|x_{i,j})$, which expresses the probability that the pixel $x_{i,j}$ is originated from a surface of type ω_c. The $P(\omega_c|x_{i,j})$ are then iteratively updated using the information from the neighborhood $\Gamma(x_{i,j})$ of the pixel, which is constituted by the pixels contained in an $l \times l$ window centered in $x_{i,j}$, except $x_{i,j}$ itself. In order to take into account the interaction between neighboring pixels, Besag [17] introduced the concept of neighborhood potential
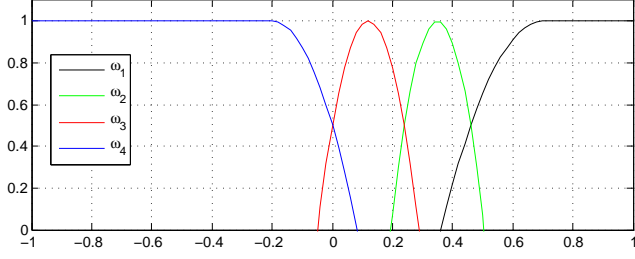
Fig. 5: Parabolic functions Λ_c, $c = 1, \ldots, 4$ for the four classes.

	cluster centers m_c		transition values b_k
ω_4	-0.2	ω_4/ω_3	0
ω_3	0.12	ω_3/ω_2	0.24
ω_2	0.35	ω_2/ω_1	0.46
ω_1	0.70		
(a)		(b)	

TABLE I: Class centers and transition values of the four detected classes.

$U(x_{i,j})$, which we compute as

$$U(x_{i,j}|\omega_c) = \sum_{x \in \Gamma(x_{i,j})} [1 - P(\omega_c|x)]. \qquad (8)$$

Then, the spatial membership $P_{spat}(x_{i,j}|\omega_c)$ probability can be computed as:

$$P_{spat}(x_{i,j}|\omega_c) = \gamma e^{-\beta U(x_{i,j}|\omega_c)} \qquad (9)$$

where β is a positive coefficient used for weighting the influence of the spatial context and γ is a normalization factor such that $\sum_{c=1}^{4} P_{spat}(x_{i,j}|\omega_c) = 1$. Anyway, it is not needed to actually compute γ since it will elide in the next equation. If β is small the influence of the spatial contest will have an important weight in the final computation of P_{spat}. The last step is to compute the combined spectral-spatial probability $P(\omega_c|x_{i,j})$, which is also the total probability assigned to a pixel $x_{i,j}$ to belong to a certain cluster ω_c

$$P(\omega_c|x_{i,j}) = \frac{P_{spec}(x_{i,j}|\omega_c) \cdot P_{spatial}(x_{i,j}|\omega_c)}{\sum_{c'} [P_{spec}(x_{i,j}|\omega_{c'}) \cdot P_{spatial}(x_{i,j}|\omega_{c'})]}$$

$$\sum_{c=1}^{n} P(\omega_c|x_{i,j}) = 1 \qquad (10)$$

Equations (8)-(10) are iteratively repeated for a fixed number of iterations (5 in our case), and the final $P(\omega_c|x_{i,j})$ can be used to create Cl following a maximum probability criterion.

V. EXPERIMENTAL RESULTS

We have tested the developed system collecting multiple videos using a handheld version of the camera during multiple experimental sessions around the campus of the Max Planck Institute for biological Cybernetics, Tübingen, Germany. All the computations to process the videos are performed online producing a classified image, disparity map and 3D point cloud at about 2Hz.

An example of the collected and classified images is shown in Fig. 6. On top, from left to right, we show the three camera inputs $RedL$, $NIRC$ and $RedR$. The scene, includes a concrete path in the center, snow on the left, some vegetation on the right and in the background. In $NIRC$, it is possible to observe the bright color of the vegetation, which means a high reflectance value in the NIR band. Instead, in the images acquired with the red filter $RedL$ and $RedR$ the vegetation appears darker. Although the snow looks bright in both bandwidth, it is brighter in the red band. Finally the concrete path shows a low reflectance in both bands.

The bottom raw shows in order the raw NDVI matrix $NDVI$, the image classified according to equation (2) and the classified image after the clusterization. In order to highlight the differences among the various classes, in the left image the NDVI values $-1 \leq NDVI \leq 1$ are scaled to the set $127 \leq NDVI \leq 255$. The snow, assuming negative NDVI values, is the darkest part, followed by the concrete, while the vegetation is relatively bright. The central frame shows the classification of the image according to equation (2), where blue, brown, green and dark green areas represent respectively snow/water, soil/concrete, dry vegetation and healthy vegetation. Although the classification performed in the central frame is already quite accurate, the right frame shows less noise and outliers.

A second example showing a closer point of view is provided in Fig. 7, whose subject is a small tree about 70cm tall. In addition to $RedL$, $NIRC$, $NDVI$ and the classified image, we show also the 3D reconstruction of the healthy vegetation present in the image, which includes the tree itself and some small areas on the left of the image. This information, together with the classification, can be particularly useful to not only identify a landmark in the environment and its bearing with respect to the camera frame, but also to produce estimates of its distance and size. The interested reader is invited to watch the videos collected during these experiments in the accompanying multimedia material.

In order to verify that the data obtained by our sensor are consistent with the NDVI values found in literature, we have built a database of materials from our experiments. In particular, for each material we have selected multiple points (≥ 100) over multiple frames of many different experiments and computed the mean and covariance values of the corresponding NDVI values. Fig. 8 shows the results of this statistical analysis. The median for each material is shown as a red horizontal line, while 50% of the collected data fits in the blue boxes, known as the inter-quartile range (IQR). The extreme values (within 1.5 times the inter-quartile range from the upper or lower quartile) are the ends of the dashed vertical lines extending from the IQRs. Points at a greater distance from the median than 1.5 times the IQR are plotted individually as red pluses and are considered potential outliers.

We have studied classes of material as water, asphalt, snow, concrete, plant stems and leaves. The results obtained are all congruent with the values obtained from literature,
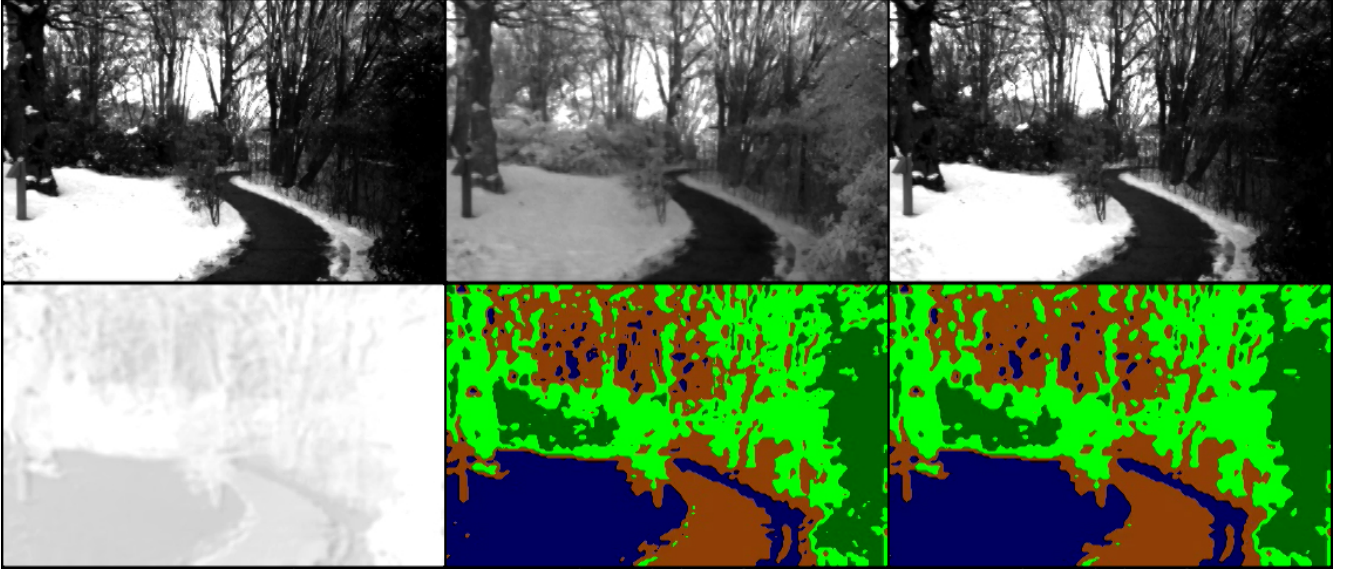
Fig. 6: Top: $RedL$, $NIRC$ and $RedR$ respectively; bottom: $NDVI$, the classification according to equation (2) and the classified image after the clusterization.
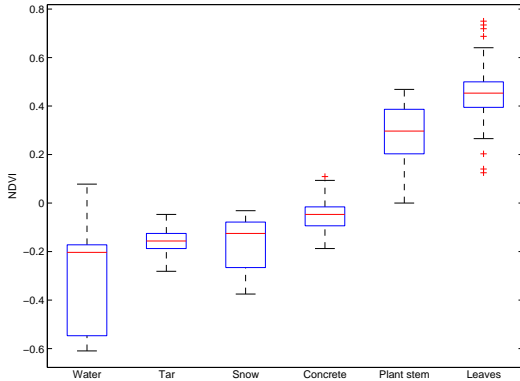

Fig. 8: Quartiles of the NDVI values taken by the camera sensor for six different materials: water, asphalt, snow, concrete, plant steam and leaves

with the notable exception of asphalt, for which we expected a slightly greater value. Nevertheless, this discrepancy may be due either to a not perfect tuning of the cameras or to the particular observed asphalt. Nevertheless, the computed values can be used in future to build a more refined classification with more classes.

VI. CONCLUSIONS

In this work, we have designed and implemented a low-cost low-weight sensor that is able to work both as spectral and stereo camera, consisting of a camera array with appropriate filters. The combination of these two sensors allows the extraction of information on the material of the observed objects and on their distance. As ultimate result, we are able to create a pointcloud of some selected features, as for example trees, that can be efficiently used, among other applications, as landmarks in outdoor navigation.

In future, we plan to optimize our software and port it to a more compact platform as Odroid XU which can be embedded on a UAV, having as final goal the processing of the information at $10Hz$. As for further development of the classification itself, we plan to test the sensor with multiple materials to create our own database and refine the number and parameters of the classes. In addition we would like to study how to exploit the depth-map also during the fuzzy classification. Finally, we plan to apply the extracted information to outdoor navigation for landmark-based navigation and identification of safe landing areas.

REFERENCES

[1] S. Maeyama, A. Ohya, and S. Yuta, "Positioning by tree detection sensor and dead reckoning for outdoor navigation of a mobile robot," in *Multisensor Fusion and Integration for Intelligent Systems, 1994. IEEE International Conference on MFI '94.*, Oct 1994, pp. 653–660.

[2] G. Avina-Cervantes, M. Devy, and A. Marin-Hernandez, "Lane extraction and tracking for robot navigation in agricultural application," in *Proc. ICAR, International Conference on Advanced Robotics, 2003*, Coimbra, Portugal, Jun. 2003, pp. 816–821.

[3] S. Zhang, W. Xiao, and L. Xie, "A novel feature extraction algorithm for outdoor mobile robot localization," in *Motion Planning*, Jun. 2008.

[4] R. Murrieta-Cid, C. Parra, and M. Devy, "Visual navigation in natural environments: From range and color data to a landmark-based model," *Autonomous Robots*, pp. 143–168, 2002.

[5] E. Todt and C. Torras, "Detecting salient cues through illumination-invariant color ratios," *Robotics and Autonomous Systems*, vol. 48, no. 2-3, pp. 111–130, 2004.

[6] E. Celaya, J.-L. Albarral, P. Jimnez, and C. Torras, "Natural landmark detection for visually-guided robot navigation," in *AI*IA 2007: Artificial Intelligence and Human-Oriented Computing*, ser. Lecture Notes in Computer Science, R. Basili and M. Pazienza, Eds. Springer Berlin Heidelberg, 2007, vol. 4733, pp. 555–566.

[7] W. Ali, F. Georgsson, and T. Hellstrom, "Visual tree detection for autonomous navigation in forest environment," in *Intelligent Vehicles Symposium, 2008 IEEE*, June 2008, pp. 560–565.

[8] Y. Lu and C. Rasmussen, "Tree trunk detection using contrast templates," in *Image Processing (ICIP), 2011 18th IEEE International Conference*, Sept 2011, pp. 1253–1256.
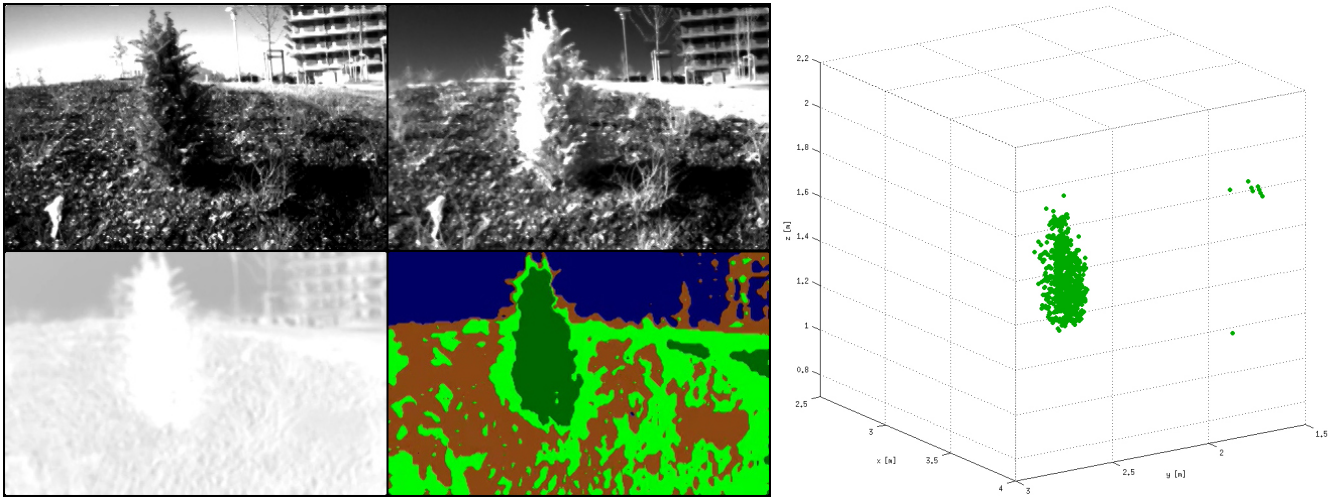
Fig. 7: Top-left: $RedL$ and $NIRC$ respectively; bottom-left: $NDVI$, the classification and the classified image after the clusterization; right: the reconstructed point-cloud of the healthy vegetation class only.

[9] M. Song, F. Sun, and K. Iagnemma, "Natural landmark extraction in cluttered forested environments," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, May 2012, pp. 4836–4843.

[10] D. Bradley, R. Unnikrishnan, and J. Bagnell, "Vegetation detection for driving in complex environments," in *Robotics and Automation, 2007 IEEE International Conference on*, April 2007, pp. 503–508.

[11] A. Kelly , A. T. Stentz, O. Amidi, M. W. Bode, D. Bradley , A. Diaz-Calderon, M. Happold, H. Herman, R. Mandelbaum, T. Pilarski, P. Rander, S. Thayer, N. M. Vallidis, and R. Warner, "Toward reliable off road autonomous vehicles operating in challenging environments," vol. 25, no. 5-6, May 2006, pp. 449–483.

[12] J. Downing, A. Murray, and A. Harvey, "Low-cost multi-spectral imaging camera array," in *Computational Optical Sensing and Imaging, International Conference*, Monterey, CA, USA, Jun. 2012.

[13] S. Labb, B. Roux, A. Bgu, V. Lebourgeois, and B. Mallavan, "An operational solution to acquire multispectral images with standard light cameras: Characterization and acquisition guidelines," in *Proceedings of International Society of Photogrammetry and Remote Sensing Workshop*, Newcastle, United-Kingdom, Sep. 2007.

[14] B. Poccia and C. Parente, "Classificazione di immagini multispettrali ikonos: Metodologie a confronto," in *Atti 12ª Conferenza Nazionale ASITA*, L'Aquila, Italy, Oct. 2008.

[15] A. M. Baldridge, S. Hook, C. Grove, and G. Rivera, "The aster spectral library version 2.0," *Remote Sensing of Environment*, vol. 113, pp. 711–715, 2009.

[16] R. Wiemker, "Unsupervised fuzzy classification of multispectral imagery using spatial-spectral features," in *Balderjahn, I., R. Mathar, and M. Schader, editors, Classification, Data Analysis and Data Highways; Proceedings of the 21. Annual Meeting of the Gesellschaft fur Klassifikation, GfKl97, Potsdam, March 1214*, Heidelberg, Germany, 1997.

[17] J. Besag, "On the statistical analysis of dirty pictures," in *Journal of the Royal Statistical Society*, 1986, pp. 259–302.