ON THE CO-REGISTRATION OF ASYNCHRONOUS MULTI-SPECTRAL AND THERMAL IMAGES

I. Cortesi[1, 2]*, A. Masiero[1], N. Pfeifer[2], G. Tucci[1]

[1] Department of Civil and Environmental Engineering, University of Florence,
Florence 50139, Italy - (irene.cortesi, andrea.masiero, grazia.tucci)@unifi.it
[2] Department of Geodesy and Geoinformation, TU Wien, 1040 Vienna, Austria - norbert.pfeifer@geo.tuwien.ac.at

KEY WORDS: Co-registration, Multi-spectral camera, Thermal camera, Litter detection, UAV

ABSTRACT:

Plastic pollution has a severe impact on the ecosystem, altering its natural equilibrium and causing serious health issues to both flora and fauna. Several actions have already been undertaken in order to reduce the plastic litter dispersion in the environment, both in terms of changing the human behavior, reducing the use of plastics and avoiding their dispersion, and of implementing methods for detecting and collecting the already dispersed ones. This paper focuses on the latter, and, in particular, on plastic litter detection on the fluvial environment. To this aim, an Unmanned Aerial Vehicle, provided with a multi-spectral and a thermal camera, have been used, in order to: (i) allow affordable periodic monitoring of relatively long river reaches, (ii) detect even quite small macro-plastics, based on their spectral signature. More specifically, since the cameras deployed in our data collection campaigns are not synchronized, this work aims at presenting the developed strategy for the co-registration of the acquired imagery, which results to be quite challenging given the few amount of visual features recognizable on the images acquired flying at a limited altitude over a river. The proposed methodology, which is based on the correlation maximization between multi-spectral and thermal images, provided reasonable results on the considered case study. The obtained values of normalized intersection over union of plastic areas are over 80%.

1. INTRODUCTION

Plastic pollution is one of the most discussed and infamous threats of our time. Every ecosystem on Earth is affected by this serious human-caused problem. The consistent use of this material and its incorrect disposal means that the health of various ecosystems is seriously at risk, including that of animals (and, obviously, humans)(Plastics Europe, 2022, United Nations Environment Program, 2018). In addition to consolidated human bad habits, the spread of the COVID-19 pandemic has increased this phenomenon by encouraging the use and diffusion of individually wrapped disposable devices (Peng et al., 2021).

Given the extent and severity of this threat, several approaches have been proposed by the remote sensing community during the last decade in order to detect the plastic litter spread in the environment. In particular, different tools and methods have been used over the years to detect and monitor floating plastic objects in fresh and salt water. Some researchers used in-situ visual census for their work (Galgani et al., 2013, Geraeds et al., 2019), but most used images or orthophotos obtained from satellite (Tasseron et al., 2021, Themistocleous et al., 2020, Topouzelis et al., 2020) or Unmanned Aerial Vehicle (UAV) imagery (Cortesi et al., 2022, Jakovljevic et al., 2020, Iordache et al., 2022, Cortesi et al., 2023).

The use of drones allows not only great flexibility in planning time and location of the surveys but also a wide range of sensors is available, which can improve the overall plastic detection performance. In particular, (Cortesi et al., 2022) suggests that the combination of multi-spectral (in the range of EM wavelength from 433 to 875 nm) and thermal data could be used in order to reduce false positives.

* Corresponding author

In accordance with this idea, the work presented in this paper, which is part of a project dealing with the problem of plastic litter detection, aims at generating a multi-band image that includes bands in the visible, near infrared and thermal part of the spectrum, which maps the scene as if acquired for each band with the identical camera orientation and at the identical time. Then, the set of multi-band images resulting from the developed procedure will be used in our future work in order to test and improve the plastic litter detection approach previously proposed in (Cortesi et al., 2022, Cortesi et al., 2021).

2. MATERIALS

The equipment used in this research includes the MAIA-S2 (by SAL Engineering and EOPTIS, Italy) multispectral camera and the DJI H20T thermal camera were mounted on a DJI Matrice 300 drone, as shown in Figure 1. In order to collect images in one flight over the case study area, a portion of the Ombrone river (near Grosseto, Tuscany, Figure 2). The area mapped by the individual images usually included a few elements: the river water, plastic samples (six in total, and differing in size, color, texture, and polymer), and in some cases river banks and vegetation (Figure 2).

MAIA-S2, which was set up using a rigid joint to the Matrice 300, acquired high-resolution multi-band images (1 Hz maximum acquisition rate), with bands at different wavelengths in the visible (VIS) and near-infrared (NIR) regions, through an array of 9 cameras (SAL Engineering and EOPTIS, 2018). The characteristics of the camera are shown in Table 1.

The MAIA-S2 camera can communicate with an external GNSS (RTK) receiver mounted on the UAV, with corrections provided by a GNSS base station located near ($\approx$ 100 m) the flight area.

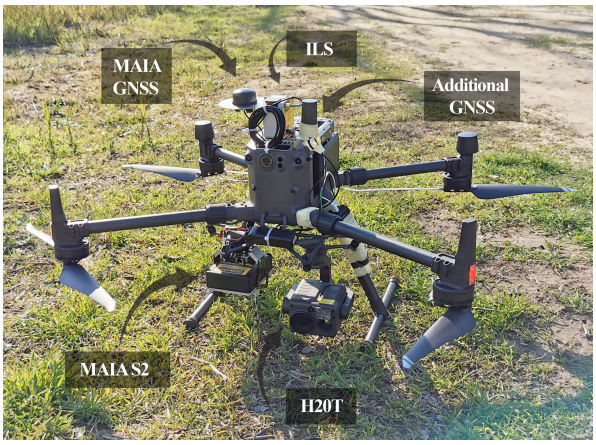


Figure 1. MAIA-S2 multispectral camera, H20T DJI thermal camera, MAIA GNSS antenna, ILS sensor and additional GNSS antenna mounted on Matrice 300 DJI.

Figure 2. The drone flying over the study area (Istia d'Ombrone, Grosseto, Italy).

The MAIA-S2 raw multi-spectral output requires radiometric and geometric corrections and co-registration. Radiometric correction was performed using the MAIA image-processing software (provided by SAL Engineering), by exploiting the measurements of the Irradiance Light Sensor (ILS), mounted on the top of the DJI Matrice 300 (see Figure 1).

The DJI H20T is a multi-sensor camera, acquiring both RGB and thermal imagery (see Table 1). RGB and thermal images are acquired in different time instants (let t_{RGB} and t_{th} be the acquisition times of an RGB and thermal image pair, then $t_{RGB} \neq t_{th}$; in both the cases 2 Hz is the maximum acquisition rate), and, unfortunately, asynchronously with respect to the MAIA-S2 acquisitions (e.g. let t_{ms} be the acquisition time of the multi-spectral image temporally closer to the previously mentioned RGB and thermal images, then $t_{ms} \neq t_{RGB}$ and $t_{ms} \neq t_{th}$). Furthermore, DJI H20T being an official DJI sensor, its gimbal (partially) compensates vibrations, enabling (close-to) nadir acquisitions. Accurate time-sync between H20T acquisitions and the Matrice 300 GNSS (RTK) receiver is not guaranteed, hence the GNSS-based exposure locations are usually determined with an error at submeter level.

MAIA-S2	
Focal length	7.5 mm
Diagonal FoV	43°
Image resolution	1208 × 960 pix
Spectral band	433 - 875 nm
Infrared Thermal Camera	
Focal length	13.5 mm
Diagonal FoV	40.6°
Image resolution	640 × 512 pix
Spectral band	8000 – 14000 nm
Wide RGB Camera	
Focal length	4.5 mm
Diagonal FoV	82.9°
Image resolution	4056 × 3040 pix
Spectral band	400 - 700 nm

Table 1. Characteristics of the MAIA-S2 and DJI H20T cameras.

The data-set initially included 654 multi-spectral images and 414 thermal images acquired at different altitudes between 10 and 80 meters above river level. Among them, 26 image couples have been selected for the initial performance investigation of the co-registration strategy proposed in this paper. The selected images have been acquired at different altitudes, varying among the 10–80 m interval. The selected image couples are formed by each thermal image and the multi-spectral one acquired at the most similar time instant, based on the timestamps provided by MAIA-S2 and DJI H20T. Since MAIA-S2 acquired images at 1 Hz, the longest possible time interval between the two acquisitions in an image couple is 0.5 s.

Plastic objects are assumed to be (almost) static. It is worth to notice that the latter is a quite unreliable assumption in general working conditions, however, it is quite realistic in this case study. The plastics were tied with fishing lines in order to allow the re-collection of such plastic samples, artificially introduced in the river just for this test. Hence, despite plastics surely moved during the data collection, their movement during the time interval between the corresponding multi-spectral and thermal image acquisitions was usually quite small.

3. METHOD

This work deals with the co-registration process of couples of multi-spectral and thermal images, in the working conditions mentioned in the previous section.

The interior orientation parameters of both the cameras (multi-spectral and thermal) are assumed to be pre-calibrated. Given the simultaneous acquisition of the multi-spectral bands, their geometric relation is assumed to be rigid, and, consequently, the overall multi-spectral camera system can be pre-calibrated.

Instead, co-registration between multi-spectral and thermal images is complicated by their variable geometric relation, caused

by (i) the asynchronous acquisition, (ii) the different connections to the UAV (rigid/non-rigid), (iii) the UAV movement.

At least two alternative options can be considered for co-registering multi-spectral and thermal images:

1. Getting all images into one reference frame using automatic bundle block adjustment of thermal and of multi-spectral images. Orthophotos can then be generated with all the bands acquired by the different cameras.
2. Working on individual images.

Option 1 was tested but showed two problems. Firstly, it can only be applied to images from large flying heights, in which enough texture (embankement) is visible. Thus the plastic in the river is only mapped to very few pixels. Secondly, the accuracy in the registration between the thermal and the multi-spectral bundle block was a few pixel. Despite the high degree of automation, this approach could therefore not be used.

Option 2 is feasible if some assumptions are met. If only the mapped water is visible, or additionally a small section of the embankment, the scene is basically flat. Thus, images of this scene can be transformed into each other using a homography. Alternatively, as an approximation, also an affine or a similarity transform can be used. The accuracy, in this case, proofed to be higher. A good knowledge of the exterior camera parameters (e.g. direct georeferencing case), or the availability of some correspondences between the two images can be exploited in order to properly estimate the transformation. The adopted procedure will be carefully described in the following.

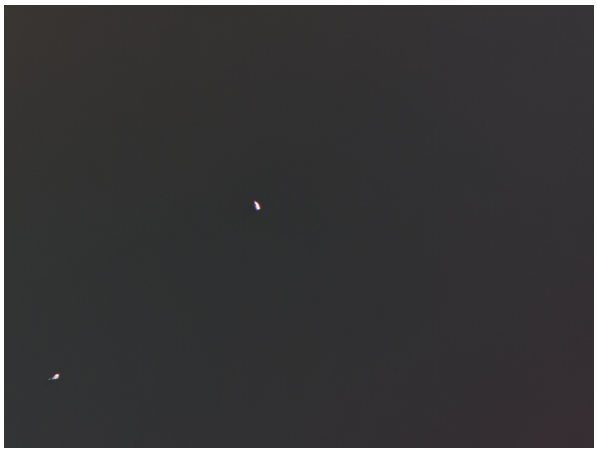
The UAV flew at low speed and at different altitudes, with most of the images acquired at quite low heights. It has previously been noted that the range of 20 m – 40 m flight height is reasonable to effectively detect plastic waste (Cortesi et al., 2022). At such altitudes, only water (and plastic litter, when present) is visible in most of the images (see Figure 3 and 4).

Since multi-spectral and thermal images were asynchronously acquired, plastic litter movements can also have a negative impact on the effectiveness of the co-registration. Nevertheless, given the working conditions described in the previous section, the plastic objects are assumed to be (almost) static during the time interval (< 0.5 s) between the multi-spectral and thermal image acquisition.

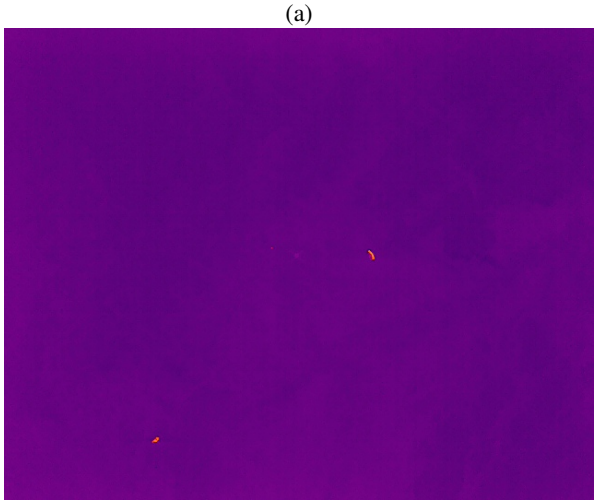
Given the characteristics of the case study, images often represent an almost planar scene. Hence, considering a thermal image and the corresponding multi-spectral one (acquired less than 0.5 s before or after the thermal one, as mentioned before), assuming an almost static scene, that distortion has already been corrected on both the two images, and that all the bands in the multi-spectral image are already co-registered, in such a way that it is possible to focus just on one of them, then there will be a homography matrix, i.e. a non-singular 3×3 matrix, such that

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \cong H \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} \tag{1}$$

where $\begin{bmatrix} u & v & 1 \end{bmatrix}^\top$ are the homogeneous coordinates of a point in the thermal image and $\begin{bmatrix} u' & v' & 1 \end{bmatrix}^\top$ the corresponding ones in the multi-spectral one. The approximate equality in (1) clearly stands up to a scaling factor.

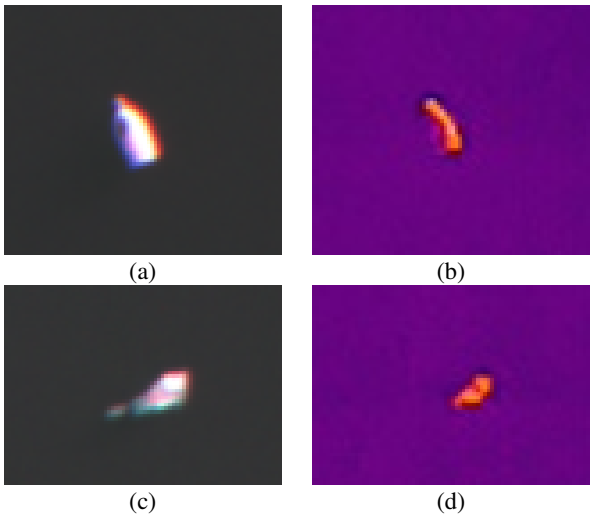


(a)



(b)

Figure 3. Example of multi-spectral (displayed Red, Green, and Blue colors) (a) and thermal images (b). One notices how a large part of the image is made up of river water



(a) (b) (c) (d)

Figure 4. Zoom of Figure 3 on the plastic samples. (a) and (c) show the multi-spectral image in RGB bands and (b) and (d) show the thermal image.

Plastic objects are usually quite distinguishable in the multi-spectral and thermal images, as shown in Figure 3 and 4. Hence, standard image segmentation methods can usually be deployed

in order to extract the plastic areas in the image, e.g. Otsu's method (Otsu, 1975). When at least four plastic objects are visible in the two images, an estimate of the homography in (1) can easily be obtained by means of the direct linear transformation (DLT) method (Hartley and Zisserman, 2003) using the plastic centroids as corresponding points. Indeed, since visual features not related to plastic objects are rarely detectable and since feature matching between multi-spectral and thermal images gave quite unreliable results, using plastic centroids currently appears as the most viable way to match points in the two images. Even not using visual descriptors, which in this case unfortunately gave unreliable results, the correspondences between plastic centroids in the two images can be determined automatically in certain working conditions, for instance: (i) if good estimates of the exterior camera parameters are available, provided by sensors on the platform (direct georeferencing case), (ii) exploiting accurate GNSS-based exposure locations and a Random Sample Consensus-like (RANSAC, (Fischler and Bolles, 1981)) procedure to determine the most reliable correspondences, (iii) a nearest neighbor approach can be used if the two camera views are very similar. Manually establishing the correspondences has to be used if the automatic methods do not work.

Actually, (i) cannot be used in the considered case, because reliable measurements of the camera attitude are not accessible. Furthermore, some issues may arise in the (ii) case as well, in particular in presence of sunglint, which may introduce many outliers in the detected regions obtained with an automatic segmentation, because the sunglint spectral signature is similar to the plastic one (Cortesi et al., 2022). RANSAC, or any of its variants, can be used in order to determine the outliers, the correspondences between the objects segmented in the two images and an estimate of the homography matrix H. Nevertheless, any RANSAC-like procedure will probably fail when the number of outliers is much larger than the number of plastic objects, as it happens in certain of the acquired images (Cortesi et al., 2022). (iii) also cannot be used, because, since the attitude of the MAIA-S2 camera is not gimbal-compensated, the two camera views can be different. Motivated by these observations, the plastic segmentation-association procedure is currently assisted by a human operator in our current implementation.

In most of acquired images the number of visible plastic objects rarely is ≥ 4. Furthermore, even when the number of objects is sufficient to determine a homography matrix $\hat{H}$ with DLT, exploiting the plastic centroids as mentioned above, it is clear that $\hat{H}$ is just an estimate of the homography between the two images, reasonably close to the optimal one, but not necessarily optimal. For this reason, a different approach has been implemented, either to improve the estimate obtained as mentioned above, or to solve the overall problem when the number of plastic objects is insufficient to apply the procedure described above.

The rationale of the implemented approach is that of determining the homography between the undistorted multi-spectral and thermal images that maximizes their correlation. Assuming that the scene is almost planar, then the homography can be written as follows (Ma et al., 2004):

$$H = K_{th}\left(R + \frac{\mathbf{t}\ \mathbf{n}^\top}{d}\right)K_{ms}^{-1} \tag{2}$$

where K_{th} and K_{ms} are the two camera matrices, which are assumed to be known, K_{ms}^{-1} is the inverse of the K_{ms} matrix, $(R, \mathbf{t})$ is the rigid (rotation, translation) transformation between the two camera reference systems, $\mathbf{n}$ is the normal of the scene plane and d is the camera-scene plane distance. Given the flight characteristics, d is usually much larger than the length of the translation vector $\mathbf{t}$, hence the contribution of $\frac{\mathbf{t}\ \mathbf{n}^\top}{d}$ is relatively small for quite high flying altitudes. Consequently,

$$H \approx H_\infty = K_{th}RK_{ms}^{-1} \tag{3}$$

where R just depends on the three relative angles between the two images, which shall be (quite small in absolute value and) optimized. In addition to H_∞, the considered transformation map between the two images is obtained by optimizing also a small translation on the image plane, leading to 5 parameters to be estimated.

The multi-spectral band with most similar appearance to the thermal one is the last MAIA-S2 NIR band, S9 (855–875) nm, hence, it is the one selected in the optimization procedure. Actually, despite being quite similar, some quite remarkable differences are unfortunately still visible. Fig. 5 compares the multispectral (a) and thermal (b) views of a plastic sample. The multispectral view in Fig. 5(a) has been obtained showing the three NIR bands of the MAIA-S2, namely: S7 (773–793) nm, S8 (784.5–899.5) nm and S9 (see (Cortesi et al., 2022) for a complete list of the spectral bands). The three MAIA-S2 bands shown in the figure are those with the appearance typically closer to the thermal one. It is worth to notice the different behavior of the plastic temperature in Fig. 5(b) with respect to that in 4(b).

(a) (b)

Figure 5. Zoom on a plastic sample: (a) three NIR bands of the multispectral image, (b) thermal image.

Motivated by the last observation, the correlation is maximized between masks of the thermal and multi-spectral images, where only plastic objects have been extracted.

To summarize, the implemented workflow is as follows:

1. Plastic segmentation on the thermal and multi-spectral images.

2. Manual establishment of the plastic object correspondences.

3. Exploit the masks of the segmented plastic object regions and the correspondences established at the previous step to optimally estimate the homography, by exploiting the reduced model (3).

4. RESULTS

The implemented approach has currently been validated on 26 couple of images, containing a total of 63 plastic object instances.

An example of the results obtained from the co-registration procedure is shown in Figure 6 and 7, which report the co-registration of the images of Figure 3.

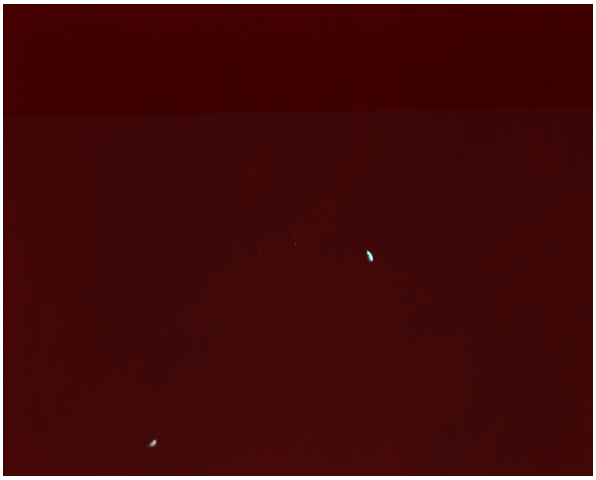


Figure 6. Co-registration result of images in Figure 3. The co-registration is shown by superimposing the thermal band with one of the NIR bands.

(a) (b)

Figure 7. Zoom of Figure 6 on the plastic samples.

Instead, Table 2, Figure 8 and 9 show the numerical results obtained on the overall validation dataset. Since in several cases the plastic appearance on the multi-spectral and thermal images is quite different, also the corresponding segmented regions have quite different areas. For this reason, table and figures show the obtained results for what concerns the Intersection over Union (IoU) between the co-registered segmented plastic regions normalized by its maximum possible value in the considered case.

To be more precise, Figure 8 shows the distribution of the normalized IoU values over the 63 plastic instances in the dataset. The corresponding numerical results (median and mean values, Median Absolute Deviation (MAD)) are reported on the first line of Table 2. Instead, Figure 9 shows the distribution of mean normalized IoU computed on each of the 26 considered images (numerical results on the second line of Table 2).

Overall, the obtained results show a quite reasonable performance of the obtained method (median normalized IoU $> 80\%$), with few plastic instances with small values of the normalized IoU, mostly when multiple plastic objects where present in the same image, and the multi-spectral/thermal appearances of the considered plastics are quite different.

	median	mean	MAD
Normalized IoU	82.2%	79.0%	17.8%
Mean normalized IoU	81.4%	79.2%	9.5%

Table 2. Co-registration results.

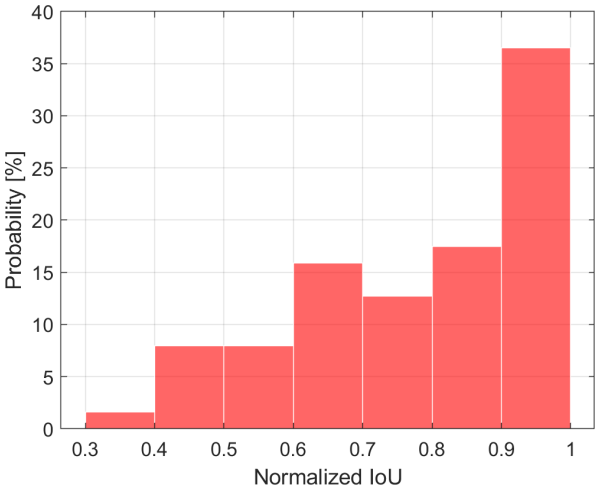


Figure 8. Distribution of the Normalized Intersection of Union values for all the plastics in the considered dataset.

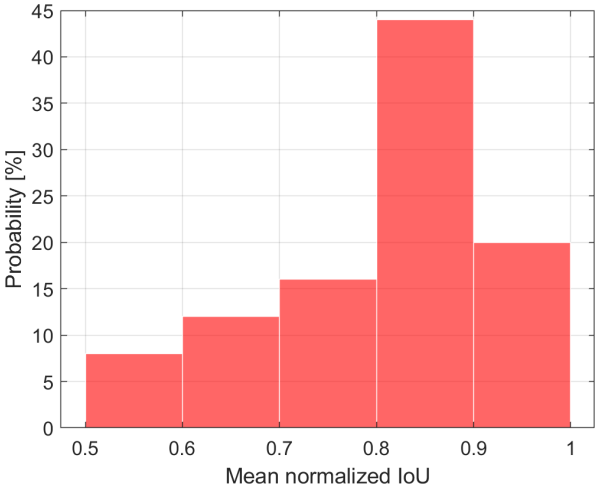


Figure 9. Distribution of the mean Normalized Intersection of Union values for all the images in the considered dataset.

5. CONCLUSIONS

The work presented in this paper is part of a project aiming at investigating plastic litter detection in fluvial environments based on their spectral signature, considering bands in the visible, near and thermal infrared. To such aim, data acquisition has been conducted using two cameras, namely the multi-spectral MAIA-S2 and thermal DJI H20T camera, mounted on a UAV flying over the case study area. Unfortunately, the two cameras acquired images asynchronously, and their geometric relation was not constant.

The presence of few visual features, mostly related to the plastic

objects themselves, in the images acquired in the considered working conditions led the authors to implement an image-to-image co-registration process, based on the maximization of the correlation between multi-spectral and thermal images.

The initial results, obtained in a validation dataset of 26 image pairs, demonstrate a reasonable performance of the method, as indicated by the median normalized Intersection over Union value greater than 80%.

Future developments will be dedicated to the validation of the method on a much larger dataset, currently already comprising over 300 multi-spectral and thermal image pairs to be analyzed. Furthermore, Deep Learning-based techniques will also be tested in order to find reliable matches between multi-spectral and thermal images (Sarlin et al., 2020). Tracking moving plastics over time will be investigated as well.

REFERENCES

Cortesi, I., Masiero, A., De Giglio, M., Tucci, G., Dubbini, M., 2021. Random Forest-Based River Plastic Detection with a Handheld Multispectral Camera. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 9–14.

Cortesi, I., Masiero, A., Tucci, G., Topouzelis, K., 2022. UAV-based river plastic detection with a multispectral camera. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B3-2022, 855–861.

Cortesi, I., Mugnai, F., Angelini, R., Masiero, A., 2023. Mini Uav-Based Litter Detection on River Banks. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10, 117–122.

Fischler, M., Bolles, R., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381–395.

Galgani, F., Hanke, G., Werner, S., De Vrees, L., 2013. Marine litter within the European marine strategy framework directive. *ICES Journal of marine Science*, 70(6), 1055–1064.

Geraeds, M., van Emmerik, T., de Vries, R., bin Ab Razak, M. S., 2019. Riverine plastic litter monitoring using unmanned aerial vehicles (UAVs). *Remote Sensing*, 11(17), 2045.

Hartley, R., Zisserman, A., 2003. *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Iordache, M.-D., De Keukelaere, L., Moelans, R., Landuyt, L., Moshtaghi, M., Corradi, P., Knaeps, E., 2022. Targeting Plastics: Machine Learning Applied to Litter Detection in Aerial Multispectral Images. *Remote Sensing*, 14(22), 5820.

Jakovljevic, G., Govedarica, M., Alvarez-Taboada, F., 2020. A deep learning model for automatic plastic mapping using unmanned aerial vehicle (UAV) data. *Remote Sensing*, 12(9), 1515.

Ma, Y., Soatto, S., Kosecka, J., Sastry, S., 2004. *An Invitation to 3-D Vision: From Images to Geometric Models, Interdisciplinary Applied Mathematics: Imaging, Vision, and Graphics, Vol. 26*. Springer.

Otsu, N., 1975. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296), 23–27.

Peng, Y., Wu, P., Schartup, A. T., Zhang, Y., 2021. Plastic waste release caused by COVID-19 and its fate in the global ocean. *Proceedings of the National Academy of Sciences*, 118(47), e2111530118.

Plastics Europe, 2022. Marine litter (12 january 2022). https://plasticseurope.org/sustainability/circularity/waste-management-prevention/marine-litter/ .

SAL Engineering and EOPTIS, 2018. Maia, the multispectral camera. (12 january 2022). https://www.spectralcam.com/maia-tech/.

Sarlin, P.-E., DeTone, D., Malisiewicz, T., Rabinovich, A., 2020. Superglue: Learning feature matching with graph neural networks. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4938–4947.

Tasseron, P., Van Emmerik, T., Peller, J., Schreyers, L., Biermann, L., 2021. Advancing Floating Macroplastic Detection from Space Using Experimental Hyperspectral Imagery. *Remote Sensing*, 13(12), 2335.

Themistocleous, K., Papoutsa, C., Michaelides, S., Hadjimitsis, D., 2020. Investigating detection of floating plastic litter from space using Sentinel-2 imagery. *Remote Sensing*, 12(16), 2648.

Topouzelis, K., Papageorgiou, D., Karagaitanakis, A., Papakonstantinou, A., Ballesteros, M. A., 2020. Plastic litter project 2019: Exploring the detection of floating plastic litter using drones and sentinel 2 satellite images. *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 6329–6332.

United Nations Environment Program, 2018. The state of plastics. world environment day outlook.