DRONE BASED NEAR REAL-TIME HUMAN DETECTION WITH GEOGRAPHIC LOCALIZATION

P.J. Baeck[1,*], N. Lewyckyj[1], B. Beusen[1], W. Horsten[1], K. Pauly[1]

[1] VITO, Remote Sensing Unit, Boeretang 200, 2400 Mol, Belgium
(pieter-jan.baeck, nicolas.lewyckyj, bart.beusen, walter.horsten, klaas.pauly)@vito.be

KEY WORDS: Drones, Geographic Information Systems (GIS), Human Detection, Photogrammetry, Deep Learning

ABSTRACT:

Detection of humans, e.g. for search and rescue operations has been enabled by the availability of compact, easy to use cameras and drones. On the other hand, aerial photogrammetry techniques for inspection applications allow for precise geographic localization and the generation of an overview orthomosaic and 3D terrain model. The proposed solution is based on nadir drone imagery and combines both deep learning and photogrammetric algorithms to detect people and position them with geographical coordinates on an overview orthomosaic and 3D terrain map. The drone image processing chain is fully automated and near real-time and therefore allows search and rescue teams to operate more efficiently in difficult to reach areas.

1. INTRODUCTION

Technologies are coming together in many of the existing remote sensing based applications, including disaster management. First of all, the emerging market of compact and easy to use drones and cameras is growing at a fast pace. Secondly, state-of-the art photogrammetric image processing techniques enable precise geographical pixel localization without using expensive GPS/IMU systems. And finally, Artificial Intelligence (AI) and Machine Learning (ML) allows us to extract relevant information from images, helps users to make decisions and improves by learning. In this study, we combine AI/ML and photogrammetric algorithms to detect humans and position them with geographical coordinates on an overview orthomosaic and Digital Elevation Model (DEM). Aim is to create a drone/camera generic, automated and performant image processing chain. This would allow search and rescue teams to operate more efficiently in difficult to reach areas.

Related work mainly has a focus on the human detection part and calculates the corresponding position from the drone GPS (Gaszczak et al., 2011), (Martins et al., 2016), (Bhattarai et al., 2018), (Rameesha et al., 2018). This is fine if you assume 1) accurate GPS tagging, 2) a nadir pointing camera, 3) no camera and lens distortions and 4) a flat terrain. This study innovates the way how resulting pixel coordinates of detected humans in individual images are translated to real world coordinates. A photogrammetric approach based on the camera's internal and external calibration parameters and generated point cloud is proposed.

2. DATASET

The dataset is part of the EU-funded AIRBEAM (AIRBorne information for Emergency situation Awareness and Monitoring) project, demonstrating the availability of unmanned platform solutions for security purposes. It worked on providing an integrated framework for crisis management in medium- to large-scale areas, in addition to developing accompanying technological components and assessing the potential of these services. In addition, the project worked on assisting the emerging market of civilian remotely piloted aircraft systems and convincing regulatory stakeholders that this technology is ready for widespread use (Maronne, 2015).

Aim of one of the validation exercises was to localize distressed people on an overgrown historical site surrounded by a 40 meter wide moat (Fort Broechem, Belgium, see Figure 1).

Figure 1. Single image of the Fort Broechem site

In total four human-sized dummies were distributed over the site. The area was then monitored using a fixed wing drone (senseFly eBee platform) with an optical camera.

*Corresponding author

Two flights were flown on consecutive days (see Table 1). The dummies were not moved in between flights.

ID	Date	Images
Flight 1	19-Jun-2015	232
Flight 2	20-Jun-2015	227

Table 1. Flight information

Flight altitude was 100 meter and image overlap was more than 90 %. The camera is nadir pointing, however during flight, pitch angles up to 15 degrees are not uncommon. The compact camera used was a Canon IXUS 127HS with a sensor size of 4608 x 3456 pixels. Unfortunately, the camera settings were not the best: an ISO value of 400, introducing noise in the imagery and slow shutter speeds from 1/160 to 1/500 seconds which can result in overexposure and motion blur.

3. METHODOLOGY

This study proposes a near-real time onground processing workflow that is able to detect humans from drone imagery using artificial intelligence and machine learning (Section 3.1) and to calculate its precise geographic location using photogrammetry (Section 3.2). It also allows the user to visualize the results in a GIS tool together with an overview orthomosaic and/or DEM for better situational awareness (Section 3.3).

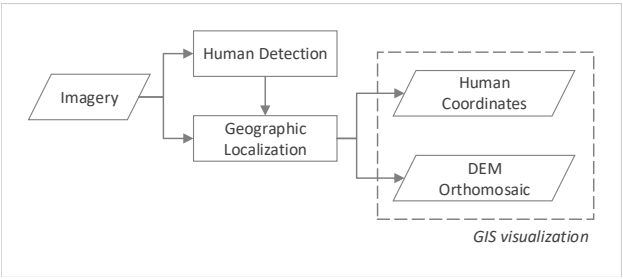


Figure 2. Image processing workflow

3.1 Human Detection

For creating a human detection model, we used the RVAI object detection software from RoboVision (RoboVision, 2019). This framework allows to go from unlabeled data to a deep learning module that is easily integrated in an image processing workflow. Flight 1 (see Table 1) was selected as reference dataset to train our model. From the 232 images, 44 contained one or more of the four human-like dummies (see Figure 3). Each of the input images are first split into 8 x 8 tiles and the tiles containing a dummy are selected to be labeled with a rectangular bounding box by the RVAI framework. Of these 54 labeled tiles, about 70 % (38) are used for training, while the remaining ones (16) are used for testing.

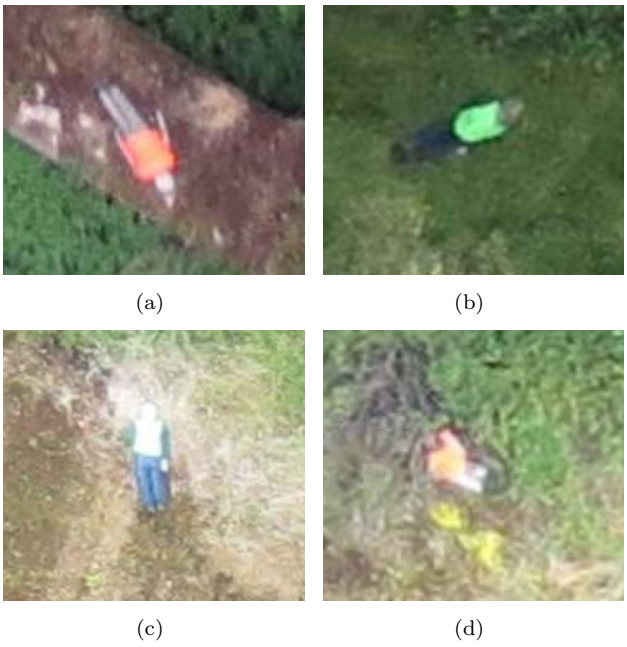


Figure 3. Four human-like dummies

The RVAI framework uses the average precision (AP) as performance metric. This defines the average of the precisions at different recall values for our dummy object class. In our case we achieved an AP of 73 %.

$$precision = \frac{TP}{TP + FP} \tag{1}$$

$$recall = \frac{TP}{TP + FN} \tag{2}$$

where TP = True Positives
FP = False Positives
FN = False Negatives

We are aware this model can be improved and made more robust by training with a much larger and diverse dataset. However this is not the main focus of this study.

The RVAI platform allows API calls from our image processing chain to evaluate the model. For each image containing a predicted dummy, the pixel coordinates of the surrounding bounding box is then stored (see Figure 4) together with a confidence factor. The confidence level threshold was set at 99.9%.

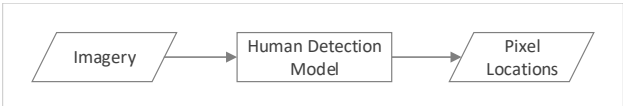


Figure 4. Human detection workflow

3.2 Geographic Localization

We make use of the Agisoft Metashape photogrammetric software (Agisoft, 2019) for calculating the camera's internal and external calibration parameters. The images

and corresponding GPS coordinates are loaded into the software. Next, the Aerial Triangulation (AT), Bundle Block Adjustment (BBA) and Point Cloud (PC) generation steps are executed with as output (see Figure 5):

- External calibration parameters: position and orientation of each image
- Internal calibration parameters: estimation of focal length, principal point and lens distortions (Brown, 1971)
- Point Cloud (PC)

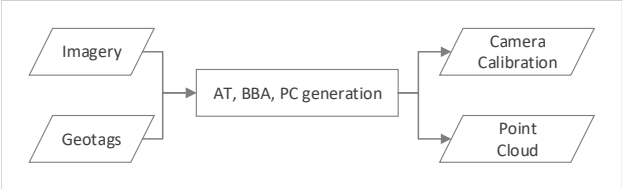


Figure 5. Camera calibration and PC generation

Again using Metashape, and based on this camera calibration model and point cloud, the pixel locations from the human detection step can be transformed to real world coordinates (see Figure 6). Instead of transforming the entire bounding box, the center pixel will be used to locate the predicted dummy.

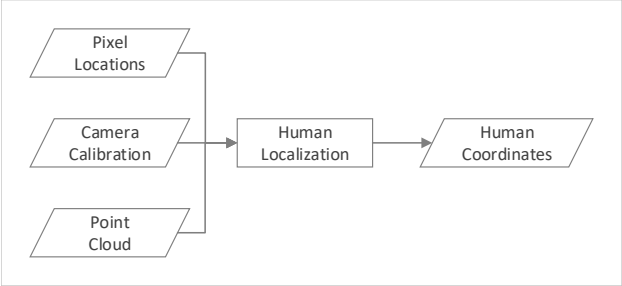


Figure 6. Human localization

3.3 GIS Integration

The Metashape software allows to calculate the DEM and orthomosaic of the region of interest (see Figure 7).

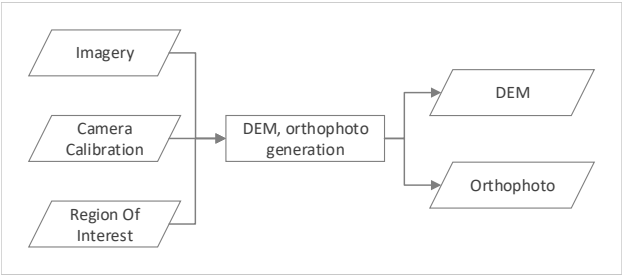


Figure 7. DEM and orthophoto generation

This information can be loaded into a GIS package such as QGIS, ArcGIS or GlobalMapper for visualizing the results. The prediction positions of human dummies are exported in Google Earth KML format and can be visualized on the orthophoto or DEM.

4. RESULTS

4.1 Human Detection Model

The 227 images of Flight 2 (see Table 1) were used to evaluate our model and were first split up in 8 x 8 x 227 = 14.528 tiles. In total 1.984 positives were detected with varying confidence level (see Figure 8).

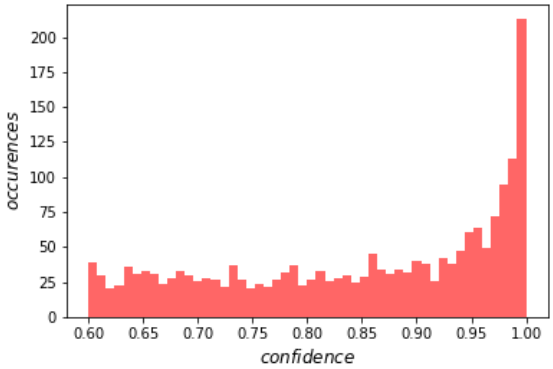


Figure 8. Confidence level histogram

The threshold was set to 99.9%, leaving just 15 positives inside the region of interest delimited by the moat of the historical site. In total 8 true positives are allocated to 4 dummies and 7 false positives are allocated to 4 other objects. All the dummies were detected, but it is clear the model does not only detect dummies. High contrast areas are common false positives (see Figure 9).

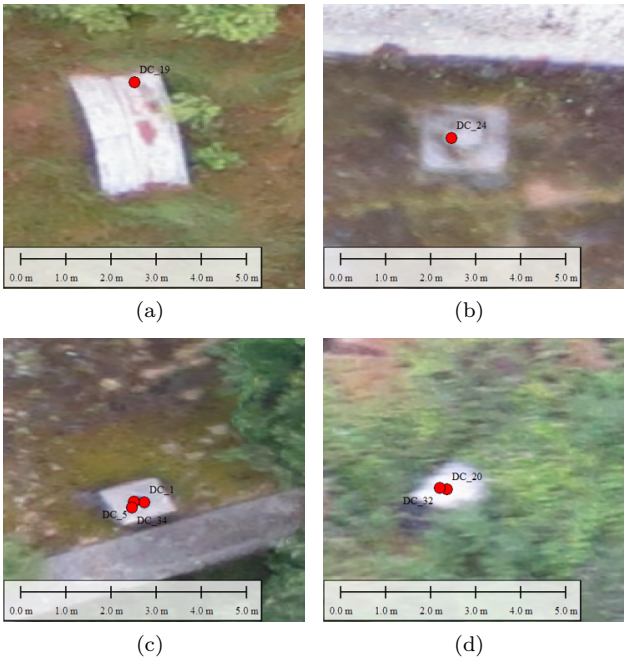


Figure 9. False Positives

We assumed our human detection classifier can be replaced with one which is better trained and more robust. However, we created our own basic detection model to proof the innovative parts of our image processing workflow.

4.2 Geographic Localization

The photogrammetric step runs successfully: 205 of the 227 camera stations are correctly aligned and there are no significant holes in the sparse point cloud, except for the moat (Figure 10). However, this approach does not allow for real-time processing, since all flight images need to be available in order to start processing.

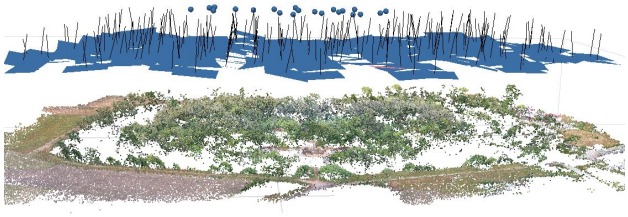
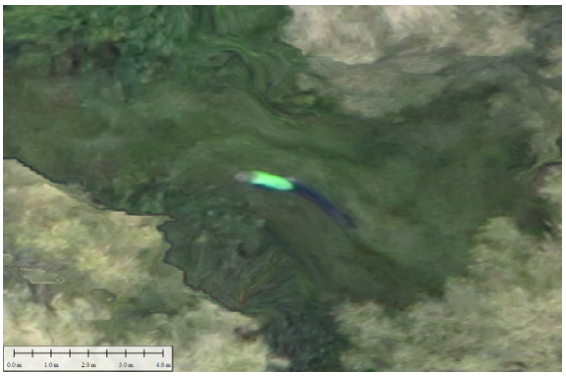


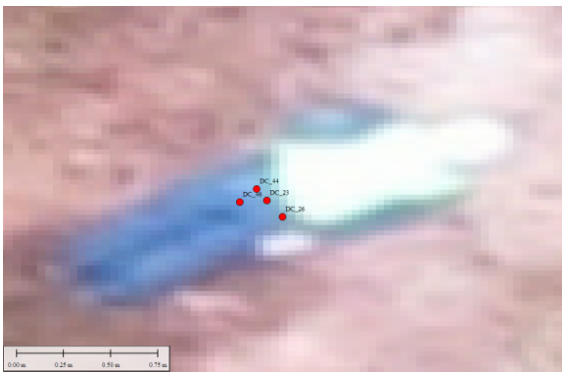
Figure 10. Camera stations and sparse point cloud

Since we have run the human detection model on the individual images and not on the final orthomosaic, we are not subjected to possible deformations present in the orthophoto due to difficult terrain, such as surrounding tree canopy (see Figure 11a).

Another advantage is that the same dummy might be visible in multiple images and we can therefore have multiple records of this same dummy in the predicted coordinates list. More hits at the same location increase the probability of dummy presence (see Figure 11b).



(a) Dummy deformation



(b) Predicted coordinates (4 hits)

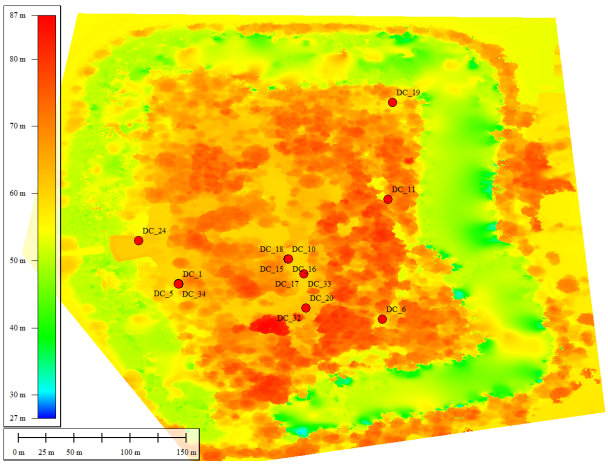
Figure 11. Orthophoto zoom on dummies

Finally, let's assume a dummy was placed under a tree or shed. With the orthophoto approach, this dummy would not be visible. However, if the dummy is visible in a single image (e.g. by flying with oblique camera angles) it is a candidate to be detected by the human classifier. This scenario was not tested due to the lack of oblique imagery.

4.3 GIS Integration

The DEM, orthophoto and predicted human locations of Flight 2 were successfully generated and visualized inside the GlobalMapper GIS tool (see Figure 12)

(a) Orthomosaic



(b) DEM

Figure 12. Flight 2 visualized with GlobalMapper. The dots represent dummy candidates.

Since the drone did not have a precise onboard GPS and we did not use Ground Control Points for absolute referencing, there will be a fixed offset on our end products. The relative planimetric accuracy is typically in the order of the size of one pixel. For Flight 2 this is 2.7 cm on average.

4.4 Preformance

The total processing time of Flight 2 on a decent GPU workstation (64 GB RAM, 16 CPU) is shown in Table 2. The human detection step runs on a dedicated server and takes about 47 seconds for a single image. It is clearly the

bottleneck in the image processing chain, while the geographic localization and GIS integration steps are making optimal use of the GPU.

Human detection	2h 40min 35sec	93.5 %
Geographic localization	5 min 07sec	3.0 %
GIS integration 2	6 min 18sec	3.5 %
	2h 52min 00sec	

Table 2. Performance of Flight 2

More work is needed in improving the run time, mainly for the human detection step.

5. CONCLUSIONS

We have presented a drone image processing chain which is able to detect human dummies in individual images, position them with geographical coordinates and visualize the results on an overview DEM and orthomosaic in a GIS environment. We are aware the detection model can be improved and made more robust by training with a much larger and diverse dataset. Also work is needed to improve the model run time.

The study innovates the way how resulting pixel coordinates in individual images are translated to real world coordinates. Since we do not work directly on the orthomosaic, this yields to better results because 1) orthomosaic deformations are avoided, 2) multiple hits of the same object increases the probability of a positive sample, and 3) oblique imagery can detect objects, which might otherwise not be visible in the orthomosaic.

This approach is independent from the type of drone (fixed wing or rotary system) or camera (RGB, multispectral or thermal) and would even allow for collaborative data gathering using an image processing solution that runs in the cloud.

REFERENCES

Agisoft, 2019. MetaShape Software, Version 1.5.1. agisoft.com (1 July 2019).

Bhattarai, N., Nakamura, T., Mozumder, C., 2018. Real Time Human Detection and Localization Using Consumer Grade Camera and Commercial UAV. *Preprints 2018, 2018110156*. doi.org/10.20944/preprints201811.0156.v1.

Brown, D. C., 1971. Close-range camera calibration. *Photogrammetric Engineering and Remote Sensing*, 8, 855–866.

Gaszczak, A., Breckon, T. P., Han, J., 2011. Real-time peope and vehicle detection from uav imagery. *Proceeding of SPIE: Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques*, 78780B–1–13.

Maronne, N., 2015. Airborne information for emergency situation awareness and monitoring. FP7-SECURITY - Specific Programme Cooperation: Security, SEC-2010.4.2-3. cordis.europa.eu/project/rcn/101536/factsheet/en (1 July 2019).

Martins, F. N., de Groot, M., Stokkel, X., Wiering, M., 2016. Human Detection and Classification of Landing Sites for Search and Rescue Drones. *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*.

Rameesha, T., Maham, R., Nimra, A., Narmeen, B., Ummay, F., 2018. DronAID : A Smart Human Detection Drone for Rescue. *15th International Conference on Smart Cities: Improving Quality of Life Using ICT and IoT (HONET-ICT)*. doi.org/10.1109/HONET.2018.8551326.

RoboVision, 2019. RoboVision Artificial Intelligence (RVAI) framework. robovision.be (1 July 2019).