

# Real time back projection onto prerecorded DHM surface of UAV recorded images with RTK GPS and IMU

Morten S. Laursen<sup>a,\*</sup>, René Larsen<sup>b</sup>, Kjeld Jensen<sup>c</sup>, Rasmus Nyholm Jørgensen<sup>a</sup>

<sup>a</sup> Department of Engineering, Aarhus University, Finlandsgade 22, 8200 Aarhus N, Denmark

<sup>b</sup> Department of AgroEcology, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark

<sup>c</sup> SDU UAS Center, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

\* Corresponding author. Email: msl@eng.au.dk

## Abstract

Within plant production drone (UAV) based aerial imaging today often relies on the creation of orthophotos. These are typically created using a mosaic of images stitched together using recognizable features within the images themselves. However, this fails when there are no features available which may be the case with tall crops, thermal images, or images taken over water.

In this work we propose a method based on using the IMU and GPS of the UAV readings together with a simple camera pinhole model of the camera lens, each pixel can be projected onto the ground surface. We are however left with an unknown variable, where does this ray strike the observed point in the scene. In this work we have chosen to use the lidar based “Danish National Height Model” (DHM). This model describes the ground height with approximately 4.5 points m<sup>-2</sup> allowing for a good approximation of the ground level. This allows us to estimate where each pixel was sampled in world coordinate space. This will enable fast positioning of areas of interest within agricultural fields.

Using blind projection based solely on the IMU the maximum error was 42.4m, whereas the mean error was 13.1m. Whereas for manual aided projection the minimum error: 0.95m, mean error: 2.9m, max error: 5.8m, standard deviation of error: 1.7m. From this we can conclude that if an observation requires higher precision than 42m a guide for a human operator to pinpoint the observation within the much narrowed window allowed for by the projection.

**Keywords:** Aerial wildlife detection, thermal imaging, image projection, aerial observation annotation

## 1. Introduction

Within plant production and related domains drone (UAV) based aerial imaging today often relies on the creation of orthophotos. These are typically created using a mosaic of images stitched together using recognizable features within the images themselves. This method is however very computer intensive and fails when there are no features available which may be the case with tall crops, thermal images, or images taken over water. It also has the downside that it is hard to perform real-time stitching as it relies on the relationship between multiple images for estimation of the terrain curvature. Bu et al. (2016) recently demonstrated a real-time approach to stitch large-scale aerial images incrementally.

The success of UAVs within plant production is relatively limited in Denmark - and currently driven by hype and expectations. A national project, RoboWeedSupport, funded in Denmark by GUDP aims at using the UAVs for close up image acquisition with a sub mm resolution and approximately one sub m<sup>2</sup> image per hectare. The images are used for semi-automated weed recognition (Dyrmann and Jørgensen 2015). In addition, grass weed is sought after in ripened cereal crops (green weed patches in yellow crops) using UAVs. Common for the latter cases is that the cost of the acquisition should not exceed approximately 1.5 EUR per hectare. Hence it is important to acquire the images in the most representative areas in the field. One solution might be to use a real time stitching of images acquired from 100 m above terrain in order to pinpoint areas in the field where close up images should be acquired.

A more complicated case is the use of thermal imaging to identify fawns in grass fields prior to swathing (Christiansen et al. 2014; Eisenbeiss, Kunz, and Ingensand 2011). In order to avoid warm artifacts due to the sun, flights must take place after sunset. The ideal conditions for spotting animals are cold clear air and windless conditions. These conditions are rarely fulfilled in Denmark in the summer. Hence ortho mosaicing e.g. using tools like Pix4D or Photoscan often fails due to a lack of features (see figure 2).

In this work we propose a method based on using the IMU and GPS of the UAV. Using these sensors together with a simple camera pinhole model of the camera lens, each pixel can be projected as a ray in 3D space describing the path on which the surface seen by the camera pixel must lie. The horizontal accuracy of the image project is evaluated based on ground truth markers. The latter is compared to visual aided repositioning of estimated markers projections compared to the automated approach.

## 2. Materials and Methods

This section describes the theory of raytracing used; the experimental setup used to acquire the thermal images of the ground truth markers, procedure repositioning the estimated location of the ground truth markers; and last estimation of

the inaccuracy of the ground truth marker positioning.

### 1.1. Raytracing

In this work we use the pinhole model (Hartley and Zisserman 2003) by which we are able to describe a camera using its intrinsic and extrinsic parameters.

The Intrinsic parameters

$$\mathbf{K} = \begin{bmatrix} \alpha_x & \gamma & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

Where  $\alpha_x$  and  $\alpha_y$  denote the horizontal and vertical focal length of the lens,  $\gamma$  the pixel skew of the sensor and  $u_0$  and  $v_0$  the principal point (intersection point between the optical axis and the image sensor, ie. where the center of projection of the lens is on the sensor).

These parameters allow us to translate a pixel position on the camera sensor into a ray in 3D in the camera's coordinate system describing from which points that pixel may have been measured.

The extrinsic parameters are a transformation matrix consisting of a rotation Matrix  $\mathbf{R}$  and a translation vector  $\mathbf{T}$ . These parameters describe how the world coordinate system is located in relation to the camera's coordinate system, thereby allowing for a translation of a camera coordinate to a world coordinate.

We rely on the ETRS89 (UTM) coordinate system as an Euclidean representation of all world coordinates.

We can calculate the rotation matrix from the photometric angles as

$$\mathbf{R} = \mathbf{R}_z \cdot \mathbf{R}_y \cdot \mathbf{R}_x$$

Where  $\mathbf{R}_z$ ,  $\mathbf{R}_y$ ,  $\mathbf{R}_x$  is the rotation about the  $z$ ,  $y$  and  $x$  axis.

$$\mathbf{R}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \pi - \Omega & -\sin \pi - \Omega \\ 0 & \sin \pi - \Omega & \cos \pi - \Omega \end{bmatrix}$$

$$\mathbf{R}_y = \begin{bmatrix} \cos \Phi & 0 & \sin \Phi \\ 0 & 1 & 0 \\ -\sin \Phi & 0 & \cos \Phi \end{bmatrix}$$

$$\mathbf{R}_z = \begin{bmatrix} \cos K & -\sin K & 0 \\ \sin K & \cos K & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Combined this allows us to describe the position of where a pixel has been sampled as

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{K}[\mathbf{R}|\mathbf{T}] \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$

Where  $x_w$ ,  $y_w$  and  $z_w$  are the world coordinates where from the pixel has been sampled,  $u$  and  $v$  is the location of the pixel in the image, i.e. pixel coordinates and  $z_c$  is our unknown scaling factor.

In order to determine the scaling factor  $z_c$  we propose to use a digital terrain model (DTM) which in Denmark has been made freely available as part of the Danish Height Model (DHM). The DTM published as part of the DHM represents the terrain height as a regularly sampled grid at 0.4m interval with a horizontal accuracy of 0.15m and a vertical accuracy of 0.05m.

$$z_{DTM} = DTM(x_w, y_w)$$

From the DTM model combined with our projection model we can define an error function simply by the discrepancy between  $z_w$  and  $z_{DTM}$  as a function of  $z_c$ .

For minimizing the error function, we define a starting guess at the origin of the camera coordinate system and a search direction towards the surface, this is to ensure that we end up with the first local minima as the first intersection with the ground in order to ensure that we cannot for example reach a point behind a hill which could define a second minima which could be chosen if using for example sea level as the starting guess.

## 1.2. Experimental setup

Images used for the demonstration of this method was captured at Nørreådal (56.435698° N 9.508677° E) using a Sensefly Ebee mounted with a ThermoMap thermal camera with a resolution of 640 x 512. The system produces tiff images containing IMU and GPS information within the exif header.

Images was taken in the period from June 14<sup>th</sup> till June 17<sup>th</sup> 2016 after sunset (10 pm - 01 am) covering approximately 4 x 37 hectares and stitched into 4 separate orthomosaics using Pix4DMapper.

In order to evaluate the performance obtained seven ground truth markers in form of aluminum foil trays (0.2 m x 0.3 m) was placed in the area and measured both horizontally and vertically using RTK GNSS.

The flights were done in darkness from two different locations indicated in figure 1. The flight altitude was set to 45 m relatively to take off point. The resulting ground pixel resolution was approximately 85 mm. Images were acquired with 50 % overlap between flight legs and approximately 90% along the flight lines.



Figure 1: Example of the four flights covering area 1 to 4 equal to approximately 150 hectares.  illustrates the seven ground markers used for ground truth. One of the ground markers (yellow) was used in two different dates

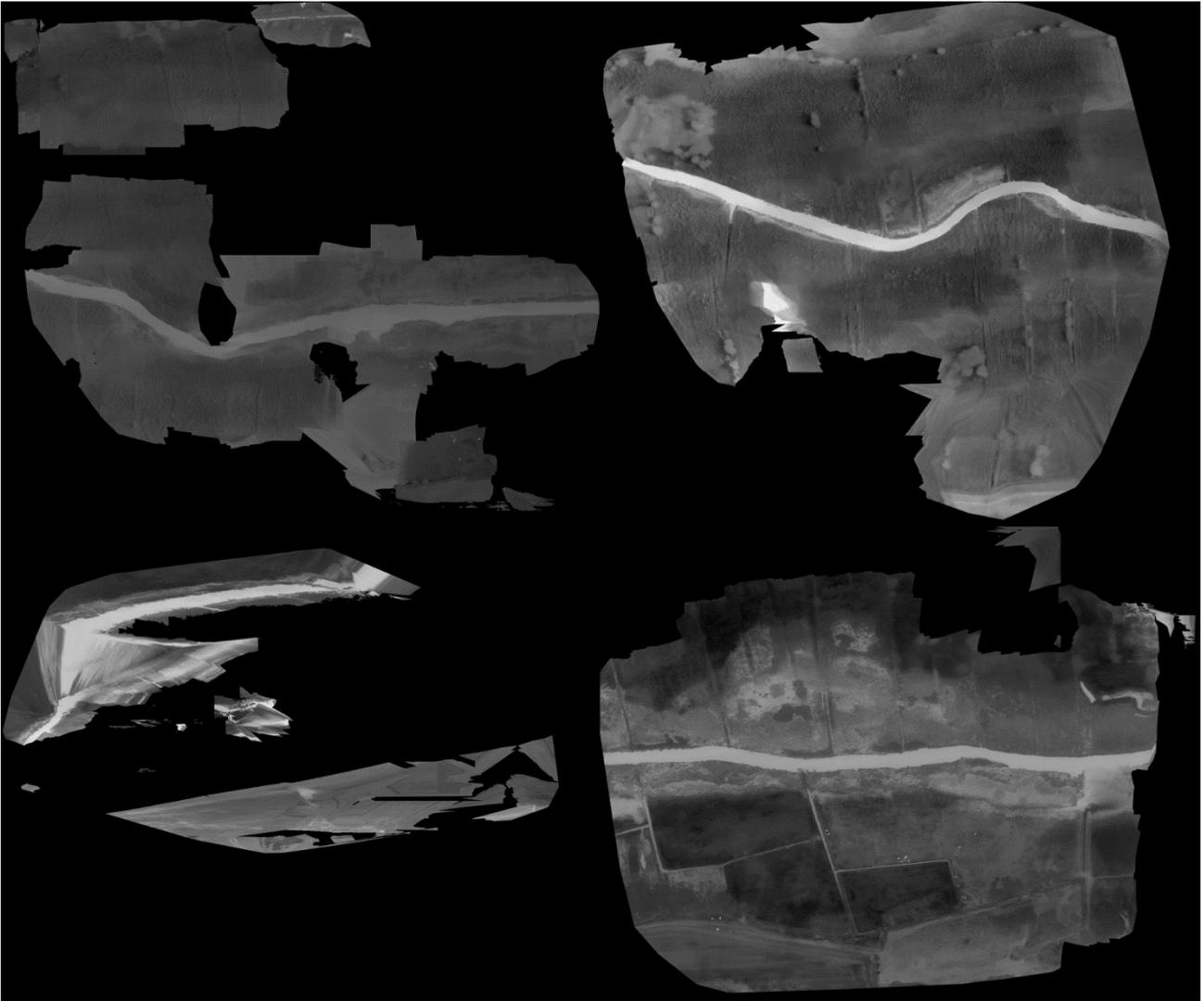


Figure 2: Illustration of the quality of the orthomosaics stitched by Pix4D for day 2, area ID 2 and 3 in the top and day 4, area ID 4 at the bottom left and, day 5 area ID 1 bottom right.

### 1.3. Orientation correction

In order to rotate from the drones navigation coordinate system (Yaw, Pitch, Roll) into photogrammetry coordinate system (Omega, Phi, Kappa) (Bäumker and Heimes 2001) we recover the local transformation from a previous dataset stitched by Pix4D, which can be found as the rotation between the IMU measurements of the images used for the stitching compared their photogrammetry orientation i.e. orientation matrix of the extrinsic camera parameters. Ensuring that all angles follows the right hand rule and flipping them as necessary otherwise.

$$R_{1ZE} = R_{IMU}^T \cdot R_{extrinsic}$$

### 1.4. Manual position estimation of ground truth markers based local features

In order to aid in the manual annotation of images such as thermal images as shown in figure 3 a user interface was written which aids the user in locating where in the world a certain photo was captured, this is performed by projecting the four image corners as described above and plotting those in the orthomosaic. The orthomosaic can be switched between a partial mosaic originating from the dataset itself and an orthomosaic automatically pulled from Kortforsyningen using WMS (OGC Web Map Service Revision Working Group 2002). The thermal image is rotated based on the yaw angle of the imu in order to align with the orthomosaic, making manual correlation simpler by not having to rotate the image mentally. In case of a single channel image such as a thermal image the image is given a false colour map. The image is annotated by repositioning the estimated marker position from the thermal image based common recognizable features in both the thermal images and the RGB orthomosaic. The current annotation point back projected onto the image in real time, while the user is moving the point in order to underline that it is the orthomosaic which has the most accurate geo information attached, within the image the contours shall therefore be manually correlated for precise positioning of the annotation.

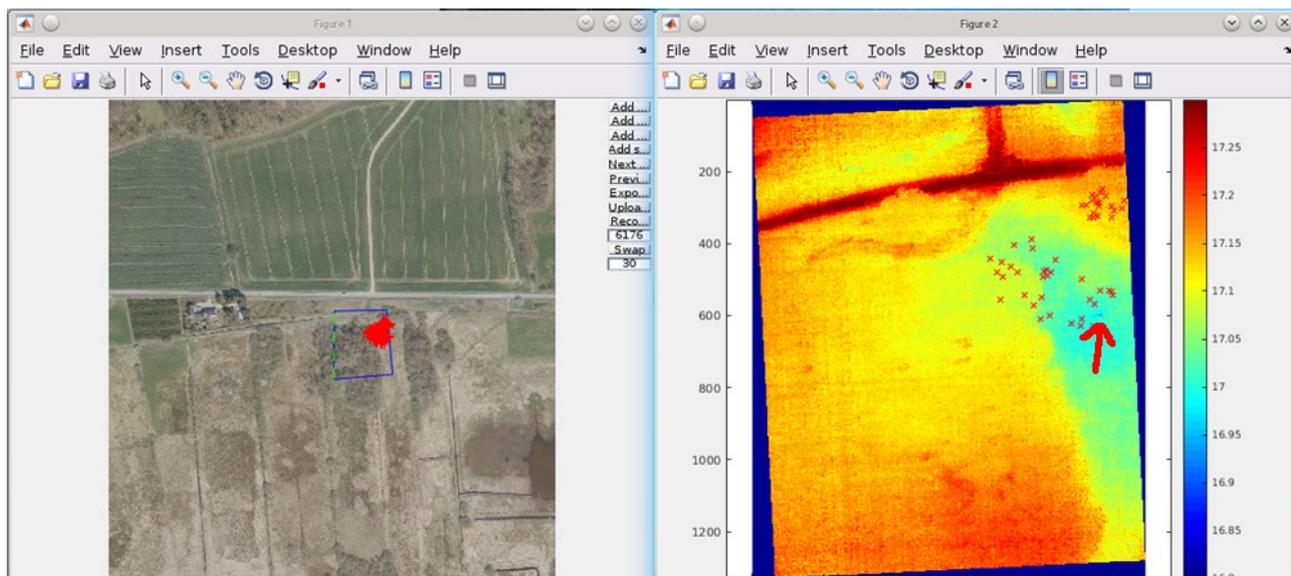


Figure 3: Example from the annotation software with expected camera field of view projected onto an orthomosaic on the left and the thermal images shown using a false colour map on the right. The red arrow indicates the exact position of the “cold” aluminum reference plate in the thermal image

### 1.5. Estimation of the inaccuracy of the ground truth marker positioning

In order to estimate the accuracy of sensor based projection each marker was annotated within the thermal camera images. For each flight two points was annotated, each in at least 10 different images. The annotated points was projected into world coordinate space and then compared RTK GNSS measured point calculating its horizontal and vertical offset and its combined distance from the point.

For estimating the performance of the manual annotation process the images in which a marker was noted was then given to a different person, which had no prior knowledge of where the markers was located. This person was then asked to estimated each marker's position using the above mentioned tool to the best of his ability. The coordinates from each estimation was then compared directly to the RTK GNSS measurements.

## 3. Results and Discussion

Manual position estimation was tested with 7 markers, one of which was used twice, to be better than 6.0m in 8/8 cases and better than 3.0m in 5/8 cases as seen in table 1. Using blind projection based solely on the IMU the maximum error was 42.4m, whereas the mean error was 13.1m. The error distribution is shown in figure 4.

During the annotation phase of the markers it was noted that rapid movements of the camera such as coming from wind gusts seemed not to be perfectly in synchronization between the IMU and what was observed within the image, one would often see a blurred image and just afterwards the model would point at a much changed pose, as the pose is extracted directly from the exif data this could indicate that the IMU sampling may not be perfectly synchronized with the camera, this could potentially be part of the cause of the large maximum error of automated positioning.

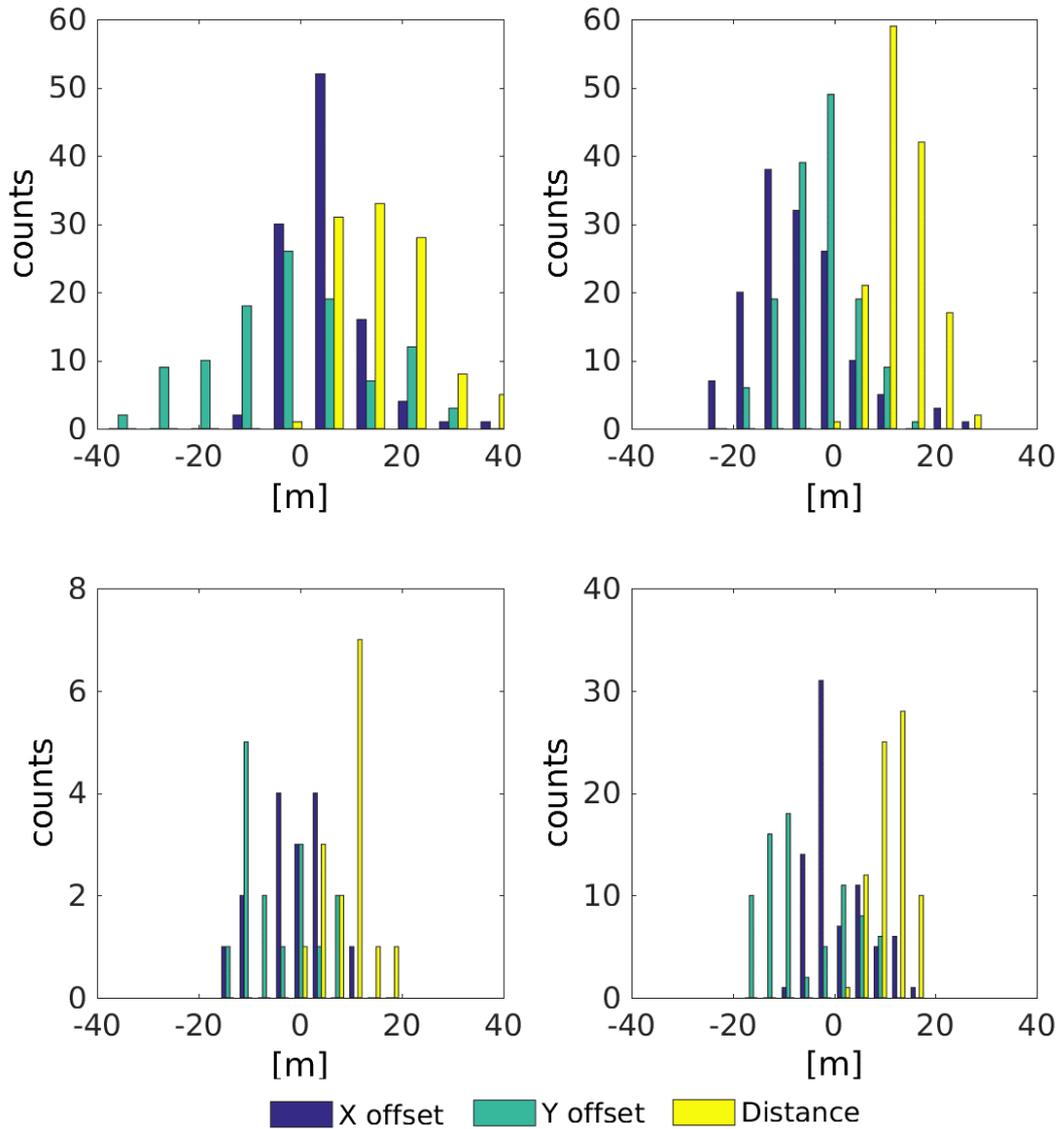


Figure 4. Error distribution of blind projection compared to reference markers, blue bars shows the offset compared to the reference between east and west and turquoise bars shows the offset between north and south, yellow bars shows distance to the reference point. Individual plots are made in order to show errors caused by fixed bias for a flight (non 0 mean of offset distance). Each subplot from top left to the right is: 14 June Area ID 2, 14 June Area ID 3, 16 June Area ID 4, and 17 June Area ID 1, respectively.

Table 1. Performance of manual annotation or relocation of ground truth markers supported by common features in the thermal image and RGB orthophoto

Date	Area ID	X offset [m]	Y offset [m]	Distance [m]	Weather
14 Jun	2	-0.7	0.6	1.0	5-6m/s
14 Jun	2	1.8	-0.9	2.0	5-6m/s
14 Jun	3	-4.0	-0.3	4.0	5-6m/s
14 Jun	3	5.0	-3.0	5.8	5-6m/s
16 Jun	4	-2.6	-1.2	2.9	rain
16 Jun	4	2.7	3.6	4.4	rain
17 Jun	1	1.6	-0.6	1.7	fog
17 Jun	1	1.0	-1.0	1.4	fog

\*Minimum error: 0.95m, mean error: 2.9m, max error: 5.8m, standard deviation of error: 1.7m

In order to further assist in the automated recognition of for example deer’s and their fawns the ground size is an important parameter, either this may be compensated for by using the embedded sensors alternatively one may fly the area

in such a way that the height over ground is held constant as shown in (Jensen et al. 2016).

#### 4. Conclusions

Using blind projection based solely on the IMU the maximum error was 42.4m, whereas the mean error was 13.1m. Whereas for manual aided projection the minimum error: 0.95m, mean error: 2.9m, max error: 5.8m, standard deviation of error: 1.7m. From this we can conclude that if an observation requires higher precision than 42m either one must rely on orthomosaic if it is possible to stitch otherwise a viable alternative is to use it as a guide for a human operator to pinpoint the observation within the much narrowed window allowed for by the projection.

#### Acknowledgements

This work is included in the public service consultancy carried out by Aarhus University for the Ministry of Environment and Food of Denmark. The project is called Wildlife Friendly Harvest and funded by Ministry of Environment and Food of Denmark. We also acknowledge the support with the thermal Ebee flights performed by Stephan Mølvig from COWI a/s. Furthermore, we would like to thank Carsten Riis Olesen, Head of Research, Danish Hunters' Association - Applied Wildlife Research and Simon Rosendahl Bjorholm from LMO for acquiring permits from the landowners of the respective areas.

#### References

- Bäumker, M., and F. J. Heimes. 2001. "New Calibration and Computing Method for Direct Georeferencing of Image and Scanner Data Using the Position and Angular Data of an Hybrid Inertial Navigation System." In: OEEPE Workshop, Integrated Sensor Orientation.
- Bu, Shuhui, Yong Zhao, Gang Wan, and Zhenbao Liu. 2016. "Map2DFusion: Real-Time Incremental UAV Image Mosaicing Based on Monocular SLAM." Zhaoyong. March 17.
- Christiansen, Peter, Kim Arild Steen, Rasmus Nyholm Jørgensen, and Henrik Karstoft. 2014. "Automated Detection and Recognition of Wildlife Using Thermal Cameras." *Sensors* 14 (8): 13778–93.
- Dyrmann, M., and R. N. Jørgensen. 2015. "RoboWeedSupport: Weed Recognition for Reduction of Herbicide Consumption." In *Precision agriculture'15*, 259–69. Wageningen Academic Publishers.
- Eisenbeiss, H., M. Kunz, and H. Ingsand. 2011. "A UAV-Based Roe Deer Fawn Detection System." Edited by H. Eisenbeiss, M. Kunz, and H. Ingsand. *International Archives of Photogrammetry and Remote Sensing, Proceedings of the International Conference on Unmanned Aerial Vehicle in Geomatics (UAV-g)*, Vol XXXVIII-1/C22 (November): 1–5.
- Hartley, R., and A. Zisserman. 2003. *Multiple View Geometry in Computer Vision*. Cambridge Books Online. Cambridge University Press.
- Jensen, Kjeld, Morten S. Laursen, Mathias M. Neerup, Martin Skriver, Rene Larsen, and Rasmus N. Jørgensen. 2016. "Towards UAV Contour Flight over Agricultural Fields Using RTK-GNSS and a Digital Elevation Model." In *International Conference of Agricultural Engineering*.
- OGC Web Map Service Revision Working Group. 2002. "Web Map Service Implementation Specification." 1.1.1.