NOVA
IMS
Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

REMOTE ESTIMATION OF TARGET HEIGHT USING UNMANNED AIR VEHICLES (UAVs)

Andrea Tonini

Dissertation presented as partial requirement for obtaining the Master's degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

2015

REMOTE ESTIMATION OF TARGET HEIGHT USING UNMANNED AIR VEHICLES (UAVs):

Andrea Tonini

MEGI

2015

REMOTE ESTIMATION OF TARGET HEIGHT USING UNMANNED AIR VEHICLES (UAVs):

Andrea Tonini

MGI

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

REMOTE ESTIMATION OF TARGET HEIGHT USING UNMANNED AIR VEHICLES (UAVS):

by

Andrea Tonini

Dissertation presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Business Intelligence and Knowledge Management

Advisor: Prof. Marco Painho

Co Advisor*:* Dr. Mauro Castelli

07/2019

THIS DISSERTATION IS PRESENTED IN A FROM OF MANUSCRIPT FOR SUBMISSION TO THE NTERNATIONAL PEER-REVIEWED OPEN ACCESS JOURNAL ”SENSORS”

DEDICATION

A Marta, Matias e Francesca

sensors

MDPI

Article

REMOTE ESTIMATION OF TARGET HEIGHT USING UNMANNED AIR VEHICLES (UAVs)

Andrea Tonini [1]*, Marco Painho[2] and Mauro Castelli[2]

[1] European Maritime Safety Agency, Praça Europa 4, 1249-206, Lisbon, Portugal
[2] NOVA IMS Information Management School, Campus de Campolide, 1070-312, Lisbon, Portugal
* Correspondence: M20170214@isegi.unl.pt; Tel.: +351 961 059 638

Received: date; Accepted: date; Published: date (TBD)

Abstract: estimation of target height from videos is used for several applications, such as monitoring agricultural plants growth or, within surveillance scenarios, supporting the identification of persons of interest. Several studies have been conducted in this domain but, in almost all the cases, only fixed cameras were considered. Nowadays, lightweight UAVs are often employed for remote monitoring and surveillance activities due to their mobility capacity and freedom for camera orientation. This paper focuses on how the height could be swiftly performed with a gimballed camera installed into a UAV using a pinhole camera model after camera calibration and image distortion compensation. The model is tailored for UAV applications outdoor and generalized for any camera orientations defined by Euler angles. The procedure was tested with real data collected with a regular-market lightweight quad-copter. The data collected was also used to make an uncertainty analysis associated with the estimation. Finally, since the height of a person who is not standing perfectly vertical can be derived by relationships between body parts or human face features ratio, this paper proposes to retrieve the pixel spacing measured along the vertical target, called here Vertical Sample Distance (VSD), to quickly measure vertical sub-portions of the target.

Keywords: remote surveillance, target height, UAV, pinhole model, image distortion compensation, Vertical Sample Distance (VSD)

1. Introduction

Unmanned Air Vehicles have been employed for more than two decades for military activities [1] but, nowadays, they are also widely used for civil applications. In particular, non-coaxial multi-rotors with weight below 4 kg [2] are often used to complement or, in some cases, even replace fixed video cameras for monitoring and surveillance activities [3]. In fact, UAVs can bring a very relevant added value compared with static installations: the possibility to transport and orient the camera as needed, allowing to perform pre-established survey paths or even follow a specific target, if needed [4].

Remote surveillance or monitoring activities may often require estimating the height of a target via image analysis. The target could be a tree for example, in order to monitor its growing for agricultural purposes [4], or a building, to follow contraction developments, or animals, to track cattle growing [5]. However, as we may expect, remote height estimation from image analysis is very often needed to define the exact stature of human beings. This is required to support the identification of a person of interest [6], health care purposes [7] or even for marketing [8]. There is a significant amount of studies in the literature dedicated to obtaining a person's body height from a video but, almost the totality of them considered data collected by static surveillance cameras. On the other hand, UAVs have been mainly used for estimating features' height for topographic or urban mapping using Photogrammetry and LIDAR (Laser Imaging Detection and Ranging) techniques (see for example [9], [10] and [11]). Photogrammetric techniques require having either a double camera pointing at the same target or acquiring two images from different orientations of a (static) feature. LIDAR data needs to be acquired by sophisticated devices installed in aircrafts specifically designated for this kind of survey technique. In some studies [12], human height estimation was performed with UAV using a machine learning approach. However, this approach requires a quite intensive elaboration and it cannot be always performed in near real-time.

This paper focuses on how the height of a feature standing vertically from the ground can be measured with a "regular" payload for lightweight UAVs, which is daylight Electro-Optical camera installed into steerable gimbals. The goal is to estimate the height using a single image in a swift fashion, possibly in near real-time, without the need for intense processing rapid situational evaluation and quick decision making during the UAV flight. Moreover, we need to take into the account that the UAV may operate outdoor, where topography and scene content may rapidly change and the target may be a static feature, like a tree, or dynamic, like humans or animals.

A widely used approach for height estimation from video footages requires to identify, when possible, vanishing lines in the scene (see for example [13], [14] and [15]). However, this approach has relevant setbacks: defining vanishing lines may not be always possible in an image [16] and a reference height in the scene is required to define the height of the target. Other authors have more recently proposed to estimate the height of a person standing on a floor considering a pinhole camera model after camera calibration and image distortion compensation [17]. A similar approach was also used in [18] in combination with person body height estimation using interpupillary distance, the comparison of these two methods showed that they are comparable and accurate.

It is here proposed an approach that foresees camera calibration and lens distortion correction before calculating geometrically the height of the target using the pinhole camera model. This procedure requires just a single image, or video frame, acquired with a camera fitted on-board of a lightweight UAV. The correction for lens distortion allows generating an image as it was acquired by a perfect pinhole system [19], which can be used for the mapping of a 3D scene to a 2D image. However, the correction of an entire image may be very time-consuming. The approach here described requires correcting just a very limited number of pixels, in order to reduce the elaboration time for near real-time applications. On the other hand, the camera calibration [20] requires intrinsic camera parameters, such as the focal length, and extrinsic parameters, such as camera position and orientation. This paper analyzes how these parameters can be defined when dealing with UAVs, for example the position of the camera is given by positional systems, like GPS.

The procedure was tested with real data collected with a regular-market lightweight quad-copter. A measuring pole of known length standing vertically from the ground was used as a target for the acquisition of several still images taken from different positions. For each shot, the height of the target was calculated considering the procedure described above and compared with the real height of the pole to assess the accuracy of the estimations. An analysis of the uncertainty was conducted to analyze how the error associated with the camera-to-target distance can influence the accuracy of the estimation.

The last part of this paper focuses on how estimating the vertical length of the target's subparts, which is particularly useful to define the exact human body height. In fact, the exact human stature can be estimated in a video if the subject is standing vertically from the ground in a fully straight pose. If the person has a different pose, such as standing relaxed with feet further apart and weight on both feet standing relaxed with weight on one leg, we would manage to estimate just the height of the body in that specific pose, see [6] and [21], not the real stature of the subject. In literature is well known the relationship between the height of a person its body parts obtained via experimental measures [22]. Therefore, the height a person who is not standing perfectly vertical can be derived by relationships between body parts or and human face features ratio [23] face of the person is well visible in the scene. It is here proposed to use the pixel spacing measured along the vertical target to quickly estimate the length of body parts or face portions. The spacing in the vertical direction is here called Vertical Sample Distance (VSD), which can be calculated as the GSD (Ground Sample Distance, [24]) but perpendicular to the ground (vertical axis).

2. Methods

The first part of this section describes the basic principles of the pinhole model for computer vision and processes for lens distortion compensation. After that, it is analyzed how computer vision can be performed when dealing with cameras installed into UAVs. The last part of this section presents and describe the method to estimate the height from still images or video frames acquired with cameras installed into UAVs.

2.1. Pinhole camera model and computer vision

In computer vision, cameras are usually modelled with the pinhole camera model [28]. The model is inspired by the simplest camera, where the light from an object enters through a small hole (the pinhole). This model considers a central projection, using the optical center of the camera and an image plane (that is perpendicular to the camera's optical axis, see **Figure *1***). In the physical camera, a mirror image is formed behind the camera center but, often, the image plane is represented in front of the camera center. The pinhole camera model represents every 3D world point P (expressed by world coordinates x_p,y_p,z_p) with by the intersection between the image plane and the camera ray line that connects the optical center with the world point P (this intersection is called the image point, noted with I in **Figure *1***).

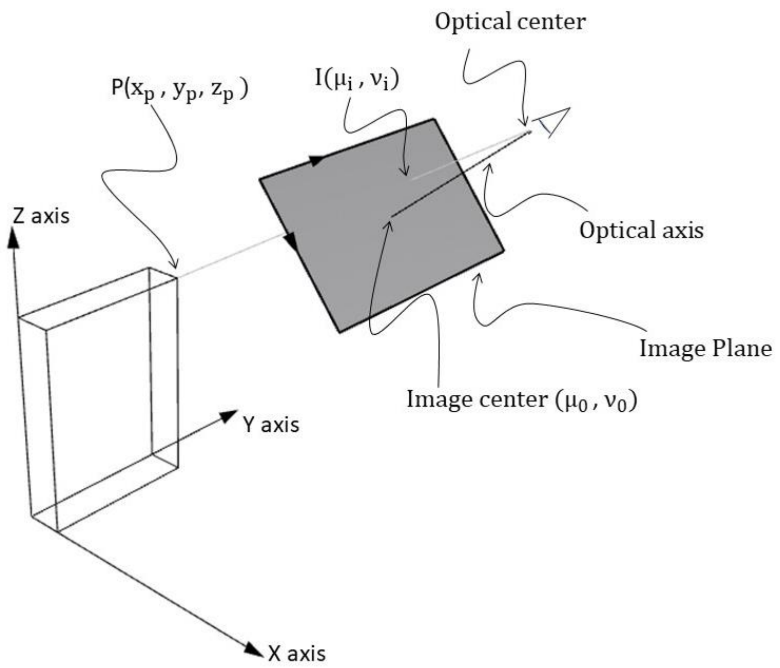


Figure 1. Graphical representation a 3D world point P is projected onto a 2D Image Plane.

The pinhole camera projection can be described by the following linear model

$$\begin{bmatrix} \mu_i \\ \nu_i \\ 1 \end{bmatrix} = K[RT] \begin{bmatrix} X_p \\ Y_p \\ Z_p \\ 1 \end{bmatrix} \quad (1)$$

Where K is the calibration matrix, defined as follow:

$$K = \begin{bmatrix} \alpha_\mu & \gamma & \mu_0 \\ 0 & \alpha_\nu & \nu_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

α_μ and α_ν represent the focal length expressed in pixels. μ_0 and ν_0 are the coordinates of the image center expressed in pixels, with origin in the upper left corner (see ***Figure 1)***. γ is the skew coefficient between the x and y axis, this latter parameter is very often 0.

The focal lengths, (which can be here considered as the distance between the image plane and optical center) can be also expressed in terms of distance (e.g. mm instead of pixels) considering the following expressions:

$$F_x = a_\mu \frac{W_\mu}{w_\mu} \quad (3)$$

$$F_y = a_\nu \frac{W_\nu}{w_\nu} \quad (4)$$

Where w_μ and w_ν are, respectively, the image (or video frame) width and length, W_μ is the width and W_ν the length of the camera sensor expressed in world units (e.g. mm). Usually, F_μ and F_ν have the same value, although they may differ due to several reasons such as flaws in the digital camera sensor or when the lens compresses a widescreen scene into a standard-sized sensor. The focal length F (assumed here for simplicity that $F = F_\nu = F_\mu$), $W_\mu$and W_ν can be also used to calculate another important element, the Field of View (FOV) of the camera, which is the angular extent of the observable world that is seen at any given moment and it may be different in μ and ν directions (see **Figure 2**).

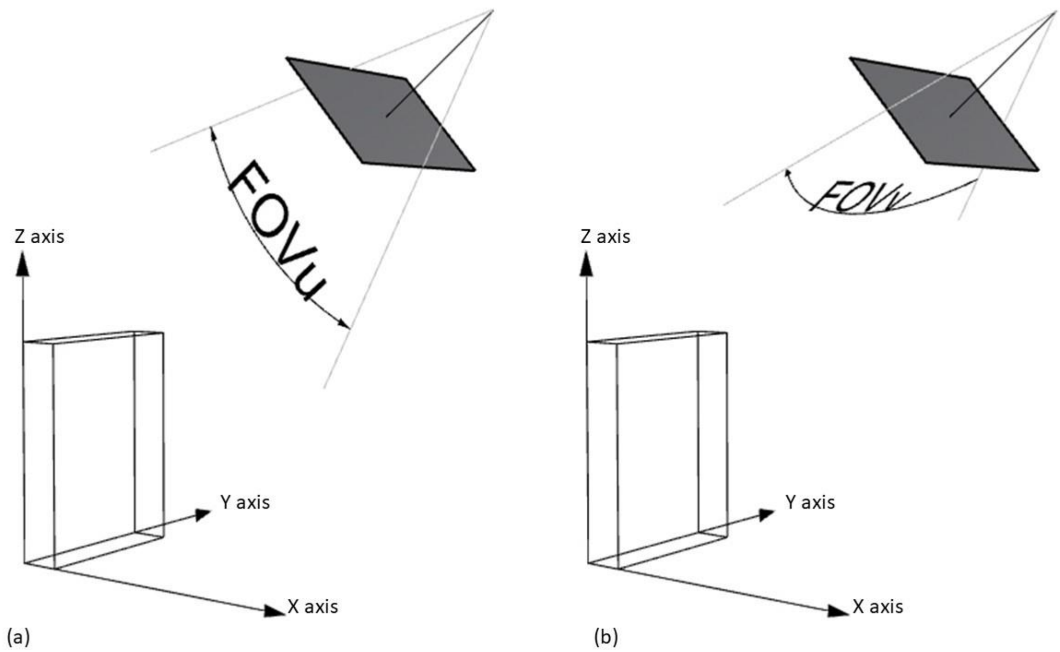


Figure 2 Graphical representation of the Field of View in the μ (a) and ν (b) directions

FOV_μ and FOV_νcan be calculated as follow:

$$FOV_\mu = 2\tan^{-1}\frac{W_\mu}{2F} \quad (5)$$

$$FOV_v = 2\tan^{-1}\frac{W_v}{2F} \quad (6)$$

R and T in (1) are the respectively rotation and translation of the camera. These are the extrinsic parameters which define the so called "camera pose".

R is defined by the axis of rotation and the angle that describes the amount of rotation. In the case of rotation around the X axis by an angle θ_x, the rotation matrix R_x is given by [19]:

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta_x) & -\sin(\theta_x) \\ 0 & \sin(\theta_x) & \cos(\theta_x) \end{bmatrix} \quad (7)$$

Rotations by θ_y and θ_z about the Y and Z axes can be written as:

$$R_y = \begin{bmatrix} \cos(\theta_y) & 0 & \sin(\theta_y) \\ 0 & 1 & 0 \\ -\sin(\theta_y) & 0 & \cos(\theta_y) \end{bmatrix} \quad (8)$$

$$R_z = \begin{bmatrix} \cos(\theta_z) & -\sin(\theta_z) & 0 \\ \sin(\theta_z) & \cos(\theta_z) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (9)$$

A rotation R about any arbitrary axis can be written in terms of successive rotations about the Z, Y and finally X axes using the matrix multiplication shown below:

$$R = R_z R_y R_x \quad (10)$$

In this formulation θ_x, θ_y and θ_z are the Euler angles.

T is expressed by a 3-dimensional vector which defines the position of the camera against the origin of the world coordinate system. GPS coordinates (Latitude, Longitude) and elevation, for example, can define T. Scaling does not take place in the definition of the camera pose. Enlarging the focal length or detector size would provide the scaling.

The next paragraph describes how the lens distortion effects and procedures for their correction.

2.2. Lens distorsion and compensation

The pinhole model does not consider that real lenses may produce several different non-linear distortions. The major defects in cameras are the radial distortion, caused by the spherical shape of the lens. Other distortions, like the tangential distortion, which is generated when the lens is not parallel to the imaging sensor, have minor relevance and will not be considered in this study. The radial distortions can usually be classified as either barrel distortions or pincushion distortions (**Figure 3**), are quadratic, meaning they increase as the square of the distance from the center.

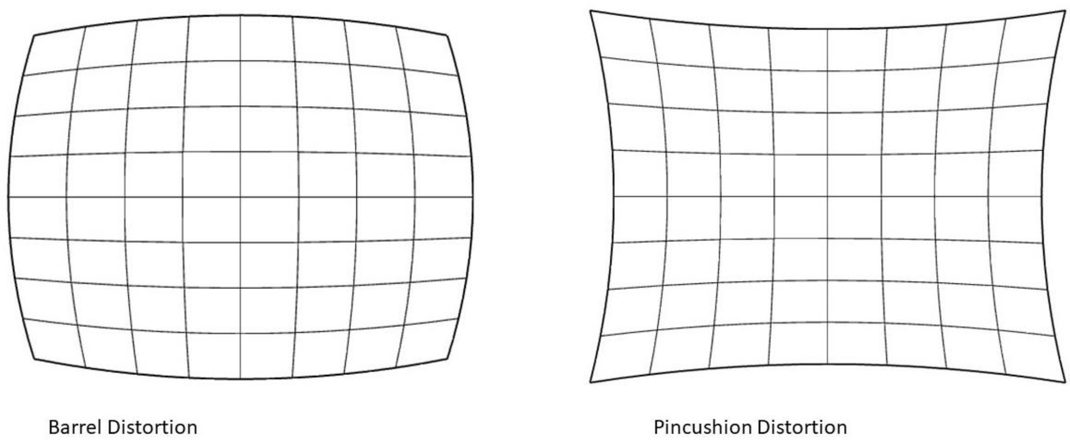


Figure 3 Effect of barrel and pincushion distortions.

Removing a distortion means obtaining an undistorted image point, which can be considered as projected by an ideal pinhole camera, from a distorted image point. The simplest way to model the radial distortion is with a shift to the pixel coordinates. The radial shift of coordinates modifies only the distance of every pixel from the image center. Let r represents the observed distance (distorted image coordinates from the center) and r_{corr} the distance of the undistorted image coordinates from the center. The observed distance for a point in the image plane I of μ_i and ν_i coordinates (see **Figure *1***) can be calculated as follow:

$$r = \sqrt{(\mu_i - \mu_0)^2 + (\nu_i - \nu_0)^2} \tag{11}$$

With these notations the function that can be used to remove lens distortion is:

$$r_{corr} = f(r) \tag{12}$$

However, before applying the compensation function $f(r)$ we need to underline that the model would be useless if images with the same distortion, but different resolutions would have different distortion parameters. Therefore, all pixels should be normalized to a dimensionless frame, where the image resolution is not important. In the dimensionless frame, the diagonal radius of the image is always 1, and the lens center is (0; 0) [25].

The formula to transform the pixel coordinates to dimensionless coordinates is the following:

$$\begin{pmatrix} p_\mu \\ p_\nu \end{pmatrix} = \begin{pmatrix} (\mu_i - \mu_0)/\sqrt{\left(\frac{w_\mu}{2}\right)^2 + \left(\frac{w_\nu}{2}\right)^2} \\ (\nu_i - \nu_0)/\sqrt{\left(\frac{w_\mu}{2}\right)^2 + \left(\frac{w_\nu}{2}\right)^2} \end{pmatrix} \tag{13}$$

Where p_μ and p_ν are the dimensionless pixel coordinates and w_μ, w_ν are the image width and height in pixels.

The dimensionless coordinates defined in (13) can be used to calculate a normalized distance r_p considering the formula given in (11). r_p can be then used to approximate the a normalized r_{corr} with its Taylor expansion [25]:

$$r_{corr} = r_p + \kappa_1 {r_p}^3 + \kappa_2 {r_p}^5 + \kappa_3 {r_p}^7 \tag{14}$$

where κ_i are the radial distortion coefficients. The "perfect" approximation would be a polynomial of infinite degree; however, this precision is not needed. Several studies, such as [26], confirmed that for average camera lenses the first order is enough, while more coefficients are required for fish-eye lenses.

r_{corr} calculated with (14) needs to be denormalized to obtain the undistorted μ_{i-corr} and ν_{i-corr} image coordinates for the image under study.

2.3. Elements to consider when dealing with cameras installed into UAVs operating outdoor

Several elements need to be taken in due consideration when operating outdoor with cameras installed into UAVs:

- The camera is usually fitted into steerable gimbals, which may have the freedom to move along one, two or even three axes (which would be formalistically called one-gimbal, two-gimbal or three-gimbal configurations, [1]). In those cases where the gimbal has limited degrees of freedom, further steering capacity for the camera must be provided by the UAV itself via flight rotations.
- The parameters required for the transformation from world coordinate system to camera coordinate (extrinsic parameters) are given by GPS measurements (latitude, longitude, and elevation) and Euler angles (yaw, pitch, and roll). Regular GPS receivers, which are not subject of enhancements such as Differential GPS, may be affected by a relevant positional error, especially in elevation. On the other hand, the orientation angles are measured by sensitive gyroscopes, which usually have very good accuracy [27].
- The parameters for the projective transformation from the 3-D camera's coordinates into the 2-D image coordinates (intrinsic parameters) must be known. For those cases where the UAV camera specs are not available, the intrinsic parameters (image principal point, focal length, and skew) can be retrieved using calibration procedure provided, for example, by computer vision open libraries such as OpenCV [28].
- The UAV can orient the camera to have the target centered in the image plane. Besides being a common practice in UAV operations for tracking, is a mandatory requirement for the calculation of height.
- The camera is usually oriented in such a way to have the feature of interest centered in the image plane. Tracking algorithms [28] can be used to automatically kept the camera pointed toward the target.
- Each video frame or still image acquired by the UAV is usually accompanied by a set of camera and UAV flight information, stored as metadata. The amount of information actually stored varies from system to system. Advanced imaging equipment may provide a complete set of metadata in KLV (Key-Length-Value) format in accordance with MISB (Motion Imagery Standards Board) standards [29]. Lightweight UAV available in the regular market are not always fitted with such advanced devices but, very often, are capable to store a minimum set of metadata which includes on-board GPS coordinates, flight orientation and camera orientation.
- Advanced UAV imaging systems are also fitted with laser range finders, which are capable to measure the instantaneous camera-to-target distance and store this information as metadata. The following paragraph describes in details the pinhole model for computer vision analysis and its parameters.

2.4. Computer vision with cameras installed into UAVs

The actual camera pose of a "gimballed" optical sensor can be determined through a sequence of homogeneous matrixes defining a number of transformations [30] that can be briefly summarized as follow:

- Transformation from Inertial frame to UAV Vehicle Frame. The UAV vehicle frame is identical to the inertial frame, only translated to the UAV position. This transformation requires a translation which only depends on the UAV's GPS location and barometric altitude measurements.
- Transformation from UAV Vehicle Frame to UAV Body frame: this transformation consists of a single rotation R, based on measurements of Euler angles that define the orientation of the UAV. In aeronautics the Eeuler angles are usually expressed through the yaw (or heading), pitch and roll.
- Transformation from UAV Body to Gimbal frame (where the origin of the gimbal frame is the center of the gimbal): this requires a translation which depends on the location of the UAV's center of mass with respect to the gimbal's rotation center and a rotation to aligns the gimbal's coordinates frame with the UAV's body frame.
- Transformation from Gimbal to Camera frame (origin at the camera's center): this transformation depends on the vector that describes the location of the gimbal's rotation center relative to the camera center and it is resolved in the camera's coordinate frame. It also depends on a simple rotation that aligns the camera's coordinate frame with that of the gimbal.

Large UAVs, which are also called MALEs (Medium Altitude Long Endurance, [31]), are usually fitted with three-gimbaled advanced imaging systems and accurate positioning systems, such as differential GPS. These systems are capable to calculate all the above-mentioned transformation in real-time and embed the instantaneous camera pose, and other information such as FOV and image footprint on ground, into the acquired video stream using the KLV (Key-Length-Value) encoding protocols [29], in accordance to military standards [32].

On the other hand, non-military lightweight UAVs available in the regular market are not always fitted with advanced imaging systems and very accurate GPS. For example, the DJI Phantom 4 PRO (a widely diffused multi-rotor platform of 1.388 kg, used to collect data for the testing of the approach described in this paper, see Paragraph 3. Results). is not capable to generate KLV embedded metadata but it can generate ancillary tags in Exchangeable Image File Format (EXIF) of still images which provide, among other information, GPS position of the aircraft, aircraft orientation and camera orientation at the moment of the acquisition of the still image. DJI Phantom 4 PRO has a GPS/GLONASS positioning system [33]. The actual accuracy of this positioning system is not specifically indicated by the UAV manufacturer, but it can be roughly assumed between 1m and 3m in the condition of good satellite signal [34]. Moreover, it is necessary to underline that the accuracy in altitude of the GPS readings is much lower than the accuracy on the horizontal plane (Latitude, Longitude). The camera of this UAV has a pivoted support (one-gimbal) with a single degree of movement along the Y axis (pitch angle, see **Figure 4**). Angular values are measured with an accuracy of +/- 0.02° [33]. Although not specified in any available technical documentation but, considering the available information of this UAV, it is here assumed that the transformation employed to provide the information in the EXIF tags are the following: a) the translation defined by the GPS coordinate of the UAV body, b) rotation based on Euler angles of the body followed by c) a 1D rotation of the camera (pitch angle). Therefore, the position of the camera when dealing with DJI Phantom 4 PRO can be defined by UAV body positional location (GPS coordinates) while the orientation is given by a yaw angle defined by flight orientation, a pitch angle defined by camera orientation and a roll angle defined by flight orientation.

The camera sensor is a CMOS of 20M effective pixels with 5472 x 3648 resolution and 13.2 x 8.8mm size, lens focal length of 8.8 mm with no optical zoom and FOV of 84°.

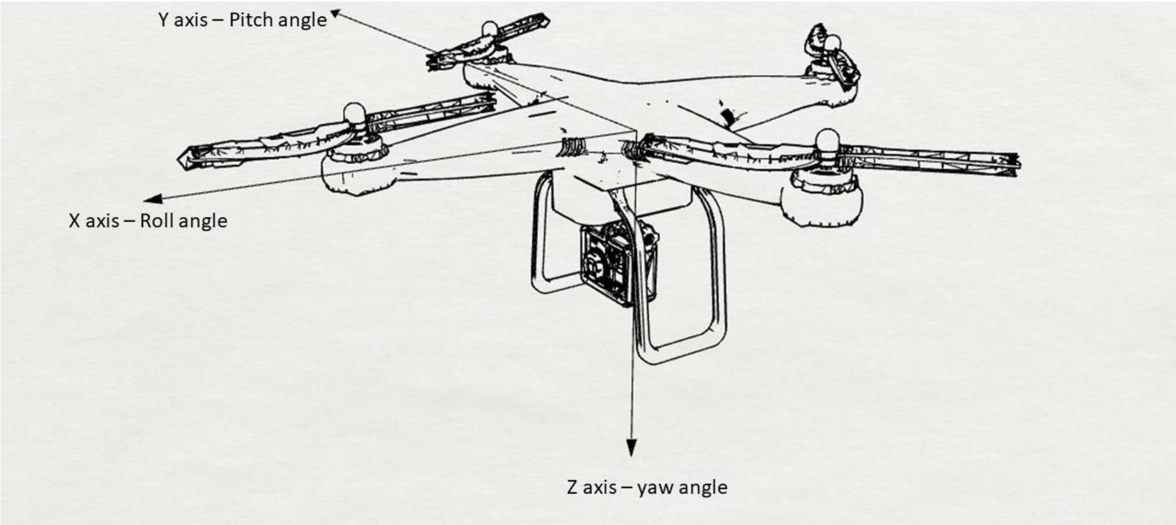


Figure 4 Axis and Euler angles for the case of DJI Phantom 4 PRO.

Let's now assume to have a lightweight UAV, like the one descripted in **Figure *4***, and a feature standing vertically on the ground, for example a pole. Let's also assume that the UAV has a heading (Yaw angle) and pitch angle appropriate to pointing to the target as graphically represented in **Figure 5**. Let's also assume that the roll angle is equal to zero 0.

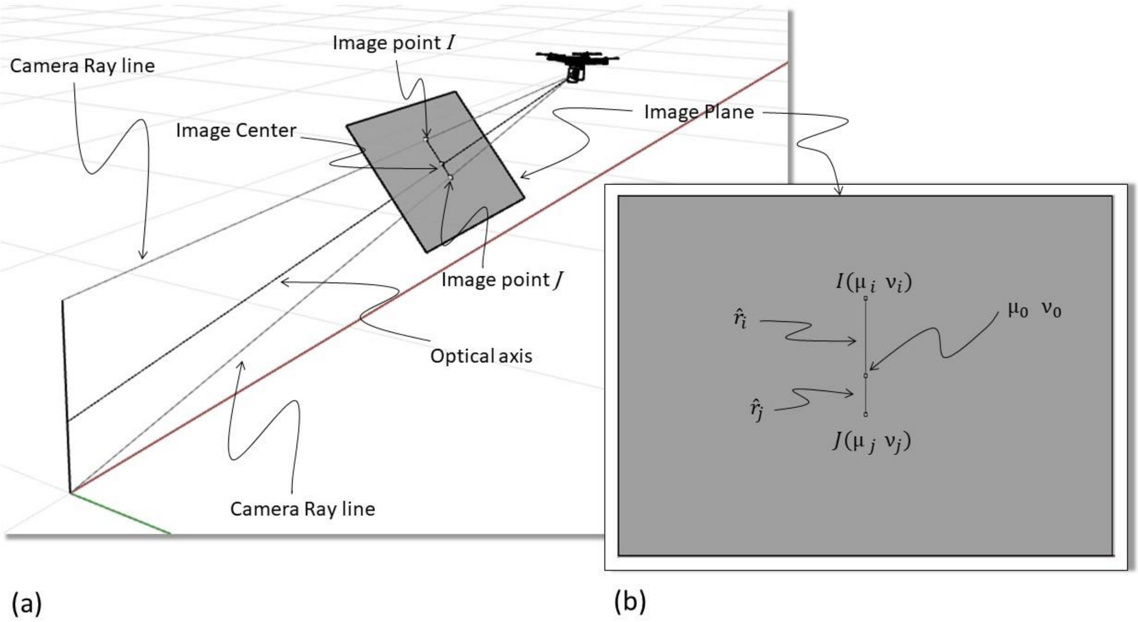


Figure 5 graphical representation of a lightweight UAV pointing to a vertical pole (a) with Roll angle equal to zero. In (b) the image plane is represented in orthogonal view (as it would appear on screen).

Point μ_0, ν_0 in **Figure 5b** is the camera center, which is obtained, as already described, by the interception between the image plane and the optical axis (see **Figure 5a**). The optical axis is centered to the target, not necessarily the midpoint but any point of the pole. The Image Point I is given by the interception of the camera ray line that connects the tip (highest point) of the pole with the camera center. This point is expressed by the image coordinate μ_i, ν_i while $\hat{r}_i$ represents the distance from the image center. Moreover, $\hat{r}_i$ is a distorted value that needs to be compensated to obtain the distance r_{i-corr} of the ideal undistorted image. The procedure to obtain such undistorted distance was already discussed in the previous paragraph (see (14). Similarly, the Image Point J is the interception of the image plane with the ray line that connects the bottom of the pole (lowest point) with the camera center. The point is expressed by the image coordinate μ_j, ν_j while $\hat{r}_j$ represents the distance from the image center that needs to be compensated to get r_{j-corr}, the undistorted distance from the center of the ideal undistorted image. The line I-J in the image plane is the height of pole expressed in pixels in the image plane.

Let's now consider the same case when the Roll angle is different than zero, as graphically represented in **Figure 6**.

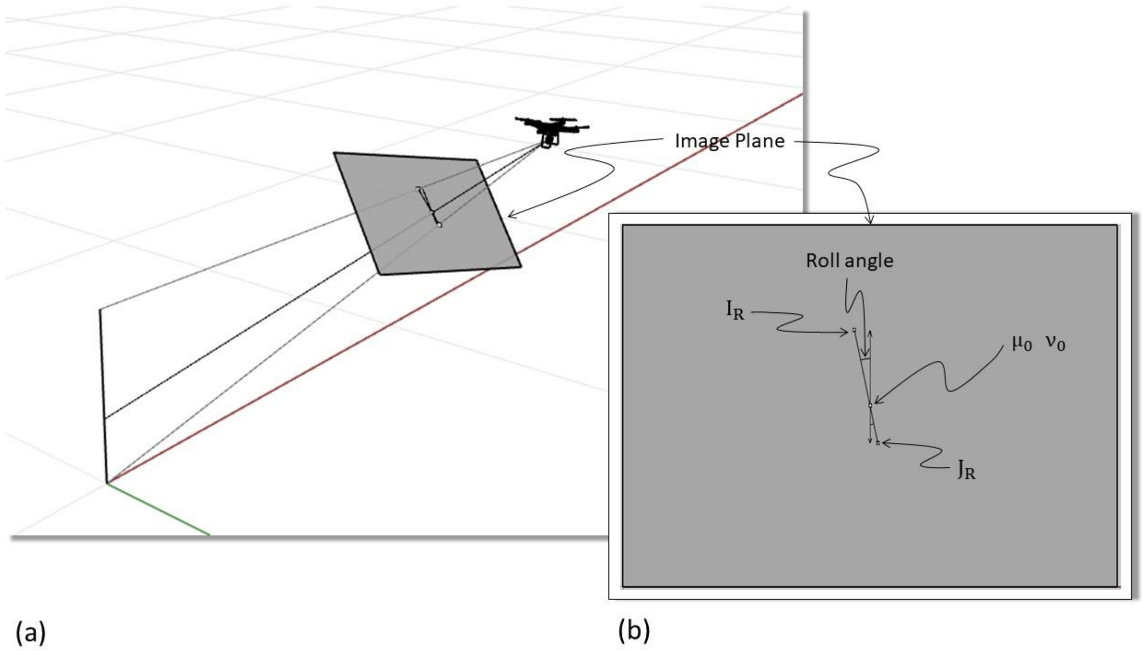


Figure 6 graphical representation of a lightweight UAV pointing to a vertical pole (a) with Roll angle different than zero. Orthogonal view of the image plane (b) with the representation of the pole and indication of the Roll angle.

When the Roll angle is different than zero, the line I_R-J_R, which is the representation of the pole in the image plane, will not appear as parallel to the ν axes, as in the case before, but rotated of an angle equal to the Roll angle itself, as it possible to infer from (7). As mentioned above, the observed distances (respectively $\hat{r}_{Ri}$ and $\hat{r}_{Rj}$) must be compensated to obtain the distances $r_{Ri-corr}$ and $r_{Rj-corr}$ of the ideal undistorted image.

The next paragraph describes how to estimate the height of a target standing vertically (pole) considering the elements discussed so far in this paper. It is used, as an example, a lightweight UAV like the DJI Phantom 4 PRO but the approach can be extended to any imaging system installed in steerable moving platforms.

2.5. Estimating target height with camera fitted into UAVs

The approach proposed in this study for the estimation of target height using camera fitted into UAVs foresees the UAV pointing at the target as depicted in **Figure** *5* and **Figure** *6*. Let's get started with the case when the roll is zero (see **Figure** *7*).

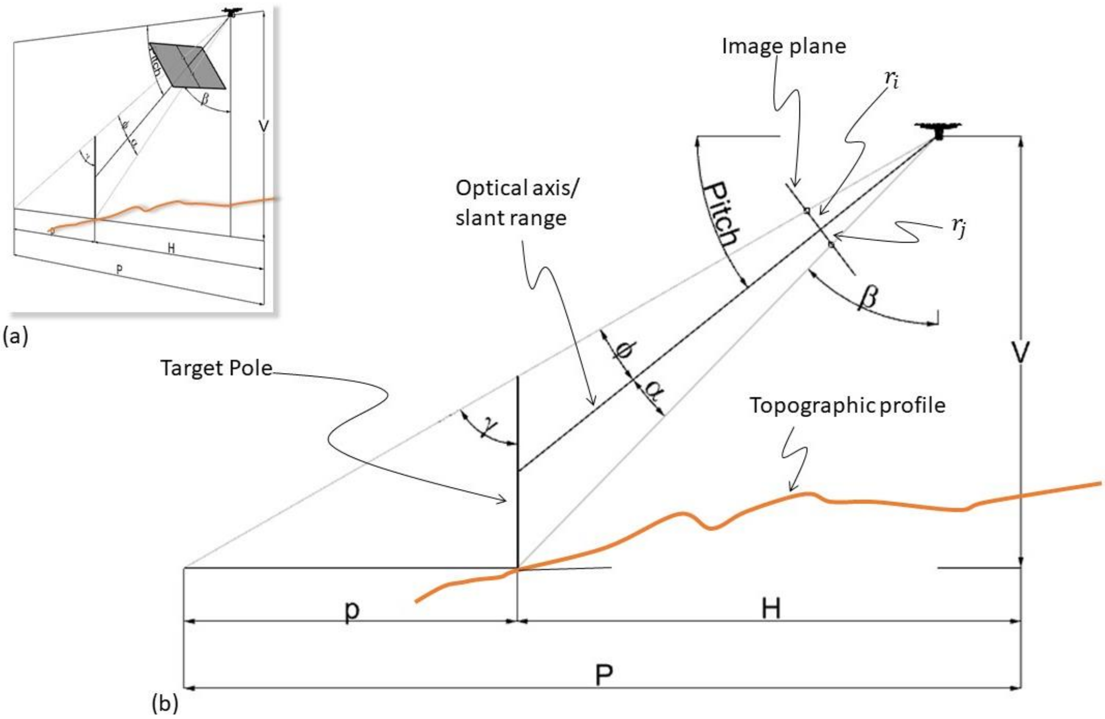


Figure 7 Perspective view of a lightweight UAV pointing to a vertical target (pole) (a). Orthogonal view of the same scene with descriptions (b).

The pitch angle, which can be also identified with θ_y, see (9) is a known value, while the angles α, β, ϕ, γ are not originally known but they can be retrieved using simple trigonometric calculations:

$$\alpha = \tan^{-1}\left(\frac{r_{j-corr}}{F}\right) \tag{15}$$

$$\beta = 90 - \left(\theta_y + \alpha\right) \tag{16}$$

$$\phi = \tan^{-1}\left(\frac{r_{i-corr}}{F}\right) \tag{17}$$

$$\gamma = (\ \phi + \ \beta + \ \alpha) \tag{18}$$

Where r_{i-corr} can be calculated considering (14) in the previous paragraph starting from the observed r_i in the image plane (see **Figure** 7). Similarly, r_{j-corr} referrers to the point J (see also **Figure** 5). F is the focal length, which was defined by (3) and (4) (assumed here for simplicity that $F = F_v = F_\mu$).

V in ***Figure 7*** is the vertical distance between the base of the target and the camera center, while H is the horizontal distance between the target and the camera center. To be underlined H and V are not related at all to the topography, as it possible to infer in **Figure 7**. If the coordinates of the target are known, then H and V are also known since the GPS coordinate of the camera are available (see Paragraph 2.3. Cameras installed into UAVs). The accuracy of the GPS and how it will impact the estimation of the target height will be treated later in this paper, but, since the accuracy of altitude in the GPS readings is much lower than the accuracy on the horizontal plane (Latitude, Longitude), V is always calculated in function of H, as defined in (19) below.

$$V = H \tan(90 - \beta) \quad (19)$$

the angles α, β, ϕ, γ are now known, as well as the pitch angle, V and H. These elements can be used to calculate the height of the target using the triangles similarity theorem. In fact, P (see **Figure 7**) can be calculated as follow:

$$P = V \tan(\alpha + \beta + \phi) \quad (20)$$

p is which is the horizontal distance between the base of the target and the camera ray that passes thought the tip (highest point) of the target, which can be calculated as follow:

$$p = P - H \quad (21)$$

Finally, the height of the target can be calculated:

$$Height\ of\ the\ Target = p \tan(90 - \gamma) \quad (22)$$

As already mentioned, the horizontal distance between the target and the camera center can be determined if the coordinates of the target are known. In practice, this could be the case only when dealing with immobile features like trees or buildings. If the position of the target is not known, as it may happen for moving targets like humans, vehicles, etc., laser range finder devices can be used to measure the instantaneous camera-to-target distance (slant range). As already mentioned, advanced imaging systems are very often fitted with such devices [1] and the instantaneous distance measurements can be stored in the KLV metadata set [31].

Slant Range values are distance is aligned to the optical axis of the camera (see **Figure 7**) and used to calculate the horizontal distance H using the following formula:

$$H = Slant\ Range\ Distance * \sin(90 - Pitch\ angle) \quad (23)$$

To be underlined that also the Slant Range distances measured by laser range finders are affected by a certain error that should duly be considered during the estimation of target height.

Let's now analyze the case when the roll angle is different than zero. In this case, as already discussed (see **Figure 6**) the points I and J are not located along the υ axis passing on the center of the image. In other words, a vertical feature will appear as "tilted" in the image on an angle equal to Roll. However, as it possible to infer from (10) and as graphically represented in **Figure 8**, I and J are in the same (vertical) plane than Pitch angle. In other words, the approach presented in this paper does not need to consider the Roll angle for the calculation of target height. Also, in this case it is necessary to perform distortion correction to obtain r_{i-corr} and r_{j-corr} and use these parameters in the formulas previously described (see (15) and (17)).

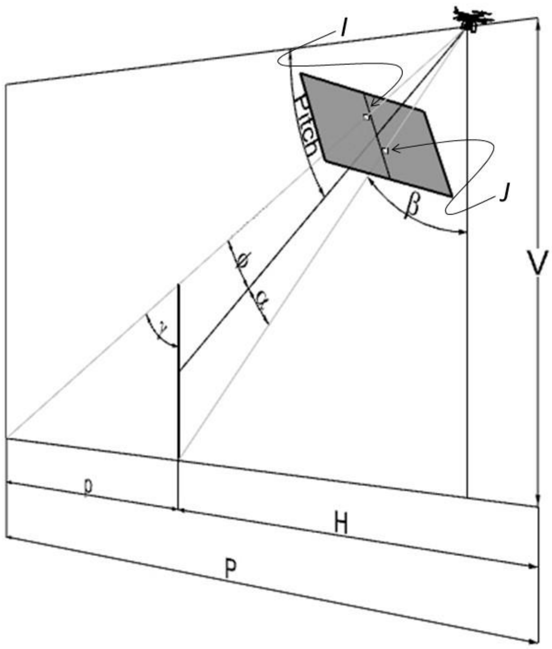


Figure 8 Perspective view of the image plane with visualization of the I and J, representing respectively the top and the bottom of the pole in the mage plane.

2.6. Workflows for target height estimation

The approach described in the following paragraphs is here summarized through workflows which are intended to be of practical use. The first workflow in **Figure *9*** should be considered during the planning phase prior to initiate a surveillance campaign to define if all the condition to estimate the height of the target feature are in place. It is necessary to underline that if the target is a moving feature and the UAV is not equipped with a laser range finder, it would not be possible to estimate the height of the target with the proposed procedure. This is relevant limitation should be addressed in future studies. The second workflow (**Figure *10***) describes the actions to perform during the UAV flight to obtain all the information needed to calculate feature height.

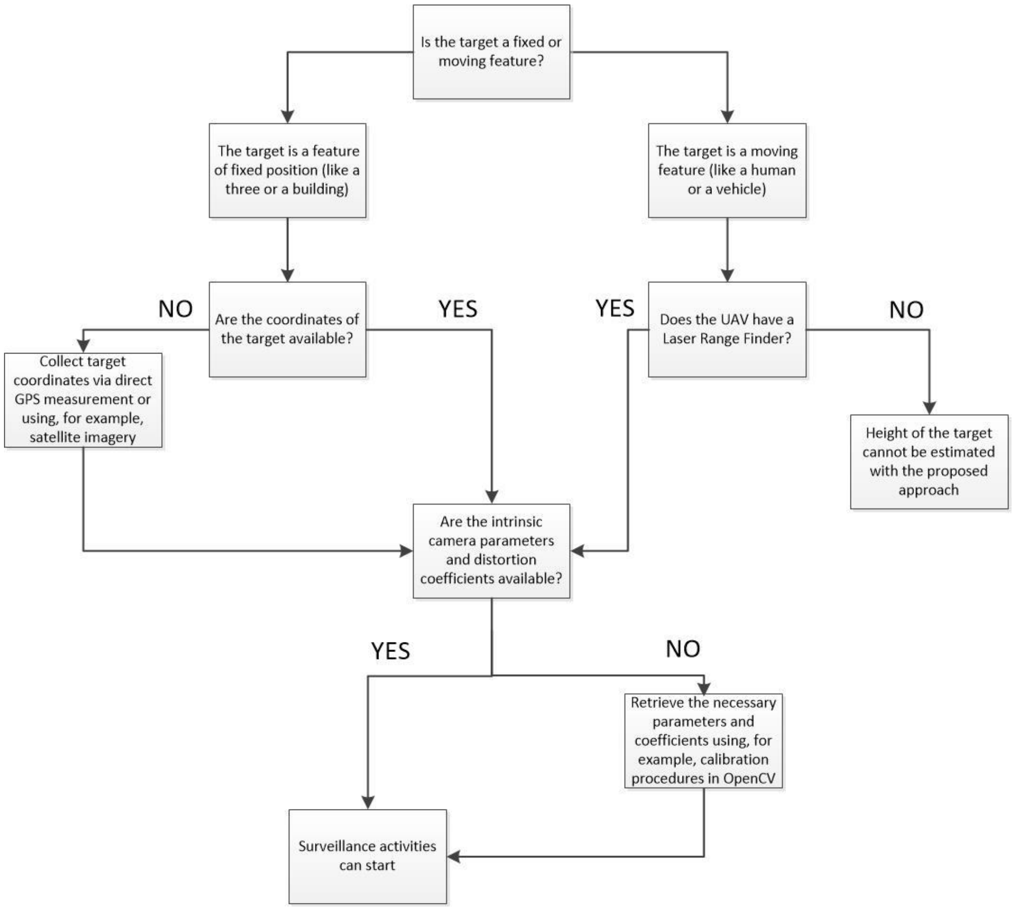


Figure 9 Workflow to verify if all the conditions to estimate the height of the target feature are in place. This analysis should be done during the planning phase prior to initiate a surveillance campaign.

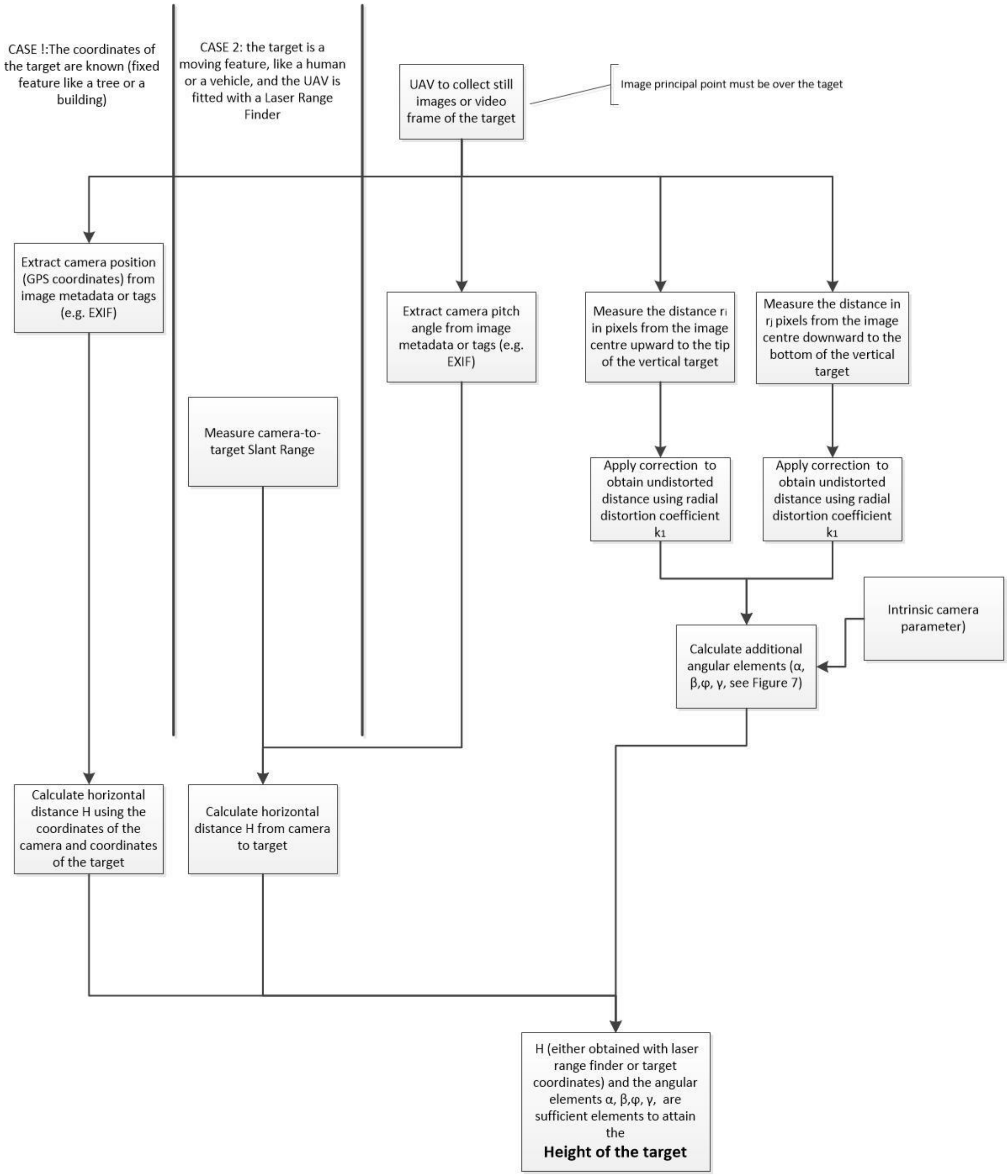


Figure 10 Workflow describing the actions to perform during the UAV flight to obtain all the information needed to calculate feature's height.

3. Results

The procedure to estimate target height described in the previous section was tested using real data acquired with a DJI Phantom 4 PRO (see Paragraph 2.3. Cameras installed into UAVs for technical details regarding this device and camera used). In this field test it was used as target a wooden pole of 180 cm standing vertically from ground located in a position of known coordinates. 32 still images were acquired with different camera poses and, in each acquired image, the principal point was always oriented over the pole (any point along pole as defined in **Figure** *6*). Images not properly oriented (principal point not located over the pole) were discarded and not used in this study. The images were acquired in an open space with good visibility to satellites.

GPS readings (position of the UAV in WGS84 geographic coordinates) and camera pitch angle of each image were extracted from EXIF tags while the number of pixels spanning upward from image principal point (r_i) and downward (r_j) were measured manually on screen (see also **Figure *11***). The Lat-Long coordinates of each image were plotted into a GIS environment along with the position of the reference pole to measure horizontal distance (H values, as graphically described in **Figure *7***). The first part of the **Table 1** and **Table 2** provides the above-discussed data for all the acquired images.

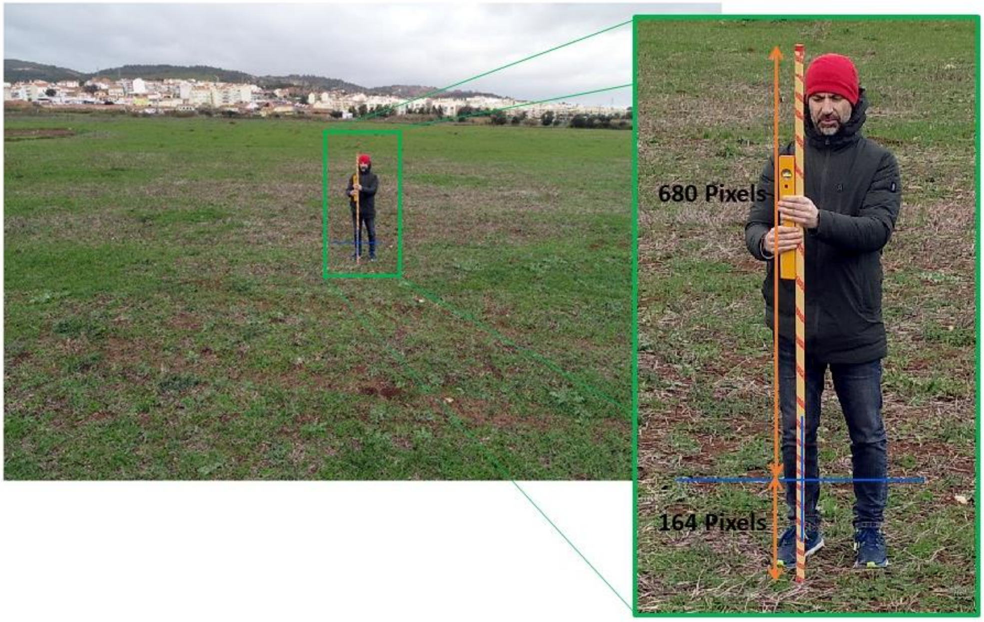


Figure 11 Still image acquired with DJI Phamtom 4 PRO with image principal point (visualized in the picture with a blue cross) located over the pole. The number of pixels spanning upward (680) and downward (164) from the image principal point were measured manually on screen. A level of 0.4m length was kept tight and alight to the pole to maintain it vertical during the acquisition of each shot.

As mentioned in the previous paragraph, the GPS readings have an accuracy between 1 to 3m. An accuracy of 1m means that the real position of the UAV is not known, but it must be located (with a probability of 95%) within a circle of 0.5 radius around the GPS readings given in the EXIF tag. Therefore, the distance of the UAV from the pole could be any value within H+0.50m and H-0.50m. In **Table *1*** and **Table *2*** this element has been reported as H+GPS Err and H-GPS Err for each image. The accuracy of the angular measurement is +/- 0.02° (see Paragraph 2.3. Cameras installed into UAVs), which is neglectable for the purpose of this study.

In the paragraph 2.2 it was described the procedure to obtain a corrected distance from image center. Such a procedure was applied to each image obtaining the r_{i-corr} (number of corrected pixels from image center upward to pole's top point) and r_{j-corr} (number of corrected pixels from image center downward to pole's bottom). These values, as well as the total number of pixels spanning the entire pole, are reported in **Table *1*** and **Table 2**. The Distortion Coefficient to be used for the correction was retrieved through camera calibration techniques [20] developed with OpenCV via Python programming.

The calculation of the target height (pole) was performed in accordance with the procedure described in the Paragraph 2.4. *Target height (NO GPS err)* in **Table *1*** and **Table 2** indicates the estimated height of the pole considering the Horizontal distance considering the camera position indicated in the EXIF tags. On the other hand, *Target height (when H+ GPS err)* and *Target height (when H- GPS err)* in **Table *1*** and **Table 2** provide the calculated height of the pole considering a GPS error of +/- 0.5m).

The field called *Height Uncertainty* in **Table *1*** and **Table *2*** represents the arithmetical difference between *Target height (when H+ GPS err)* and *Target height (when H- GPS err)*.

sensors

Table 1 data and results for the first 16 still images acquired with the lightweight UAV.

	DJI31	DJI32	DJI34	DJI41	DJ09	DJI12	DJI13	DJI15	DJI17	DJI18	DJI33	DJI35	DJI36	DJI37	DJI42	DJI_125	DJI42
Number of pixels upwards (r_i)	355	680	172	151	1358	1352	942	931	592	363	678	150	334	344	156	154	156
Number of pixels downwards (r_j)	1150	164	204	142	49	69	82	92	48	61	160	216	203	514	141	297	141
Gimbal pitch angle (degrees)	13.7	13.7	5.6	14.8	38.7	38.7	26.4	26.4	16	10.1	13.7	12	17.9	30.6	14.8	26.7	14.8
Flight roll angle (degrees)	1.8	0.9	4.2	5.2	0.4	0.8	0.5	0	0.6	0.3	1.2	2.9	3.3	0.8	1.3	3.6	1.3
Horizontal distance (H) (m)	4.2	7.9	17.1	21.28	3.6	3.64	6.04	6.17	10.28	14.96	7.95	17.38	11.62	6.04	21.12	11.78	21.12
H+ GPS err (m)	4.7	8.4	17.6	21.78	4.1	4.14	6.54	6.67	10.78	15.46	8.45	17.88	12.12	6.54	21.62	12.28	21.62
H- GPS err (m)	3.7	7.4	16.6	20.78	3.1	3.14	5.54	5.67	9.78	14.46	7.45	16.88	11.12	5.54	20.62	11.28	20.62
$r_{i\text{-}corr}$ (upward) in pixels	355	679	172	151	1352	1346	940	929	591	363	677	150	334	344	156	154	156
$r_{j\text{-}corr}$ (downward) in pixels	1146	164	204	142	49	69	82	92	48	61	160	216	203	514	141	297	141
Total number of pixels	1501	843	376	293	1401	1415	1022	1021	639	424	837	366	537	858	297	451	297
Target height (NO GPS err) (m)	1.94	1.87	1.78	1.83	1.77	1.82	1.89	1.93	1.87	1.77	1.87	1.83	1.87	1.98	1.84	1.87	1.84
Target height (when H+ GPS err)	2.17	1.99	1.84	1.88	2.01	2.07	2.05	2.09	1.96	1.83	1.99	1.88	1.95	2.14	1.88	1.94	1.88
Target height (when H- GPS err)	1.71	1.75	1.73	1.79	1.52	1.57	1.73	1.78	1.78	1.71	1.75	1.78	1.79	1.82	1.79	1.79	1.79
Uncertainty (m)	0.46	0.24	0.10	0.09	0.49	0.50	0.31	0.31	0.18	0.12	0.23	0.11	0.16	0.33	0.09	0.16	0.19

Table 2 data and results for the remaining 16 still images acquired with the lightweight UAV (continuation of **Table 1**).

	DJI143	**DJI144**	**DJI148**	**DJI149**	**DJI150**	**DJI151**	**DJI152**	**DJI153**	**DJI154**	**DJI155**	**DJI156**	**DJI157**	**DJI40**	**DJI43**	**DJI137**
Number of pixels upwards (r_i)	207	359	32	83	68	112	241	156	307	339	573	902	254	114	136
Number of pixels downwards (r_j)	208	75	86	26	102	90	145	78	304	280	559	211	180	161	33
Gimbal pitch angle (degrees)	38.5	39.4	6.9	16.6	30.4	47.3	20.7	10.7	34.4	34.4	34.5	41.7	22.9	22.4	15.5
Flight roll angle (degrees)	9.6	8.5	5.5	8.6	3.1	3	3.8	5.2	3.3	4.2	2.8	7.1	1.9	1.1	5.6
Horizontal distance (H) (m)	9.51	9.71	54.24	54.49	28.33	15.33	14.85	26.68	7.43	7.3	4.22	4.13	13.38	21.08	37.12
H+ GPS err (m)	10.01	10.21	54.74	54.99	28.83	15.83	15.35	27.18	7.93	7.8	4.72	4.63	13.88	21.58	37.62
H- GPS err (m)	9.01	9.21	53.74	53.99	27.83	14.83	14.35	26.18	6.93	6.8	3.72	3.63	12.88	20.58	36.62
$r_{i\text{-corr}}$ (upward) in pixels	207	359	29	83	68	112	241	156	307	339	573	900	254	114	136
$r_{j\text{-corr}}$ (downward) in pixels	208	75	84	26	102	90	145	78	304	280	559	211	180	161	33
Total number of pixels	415	434	118	109	170	202	386	234	611	619	1132	1111	434	275	169
Target height (NO GPS err) (m)	1.79	1.84	1.79	1.78	1.79	1.84	1.78	1.77	1.83	1.81	1.95	1.96	1.86	1.87	1.85
Target height (when H+ GPS err)	1.89	1.93	1.81	1.80	1.82	1.90	1.84	1.80	1.96	1.93	2.18	2.20	1.93	1.91	1.87
Target height (when H- GPS err)	1.70	1.75	1.77	1.77	1.75	1.78	1.72	1.74	1.71	1.69	1.72	1.73	1.79	1.82	1.82
Uncertainty (m)	0.19	0.19	0.03	0.03	0.06	0.12	0.12	0.07	0.25	0.25	0.46	0.48	0.14	0.09	0.05

Looking at *Target height (NO GPS err)* in **Table 1** and **Table 2**, the results clearly indicate that in almost no image the correct height (180cm) is obtained. However, considering *Target height when H+ GPS err* and *Target height (when H- GPS err)*, which give interval from the highest and lowest possible height value considering the GPS error, we can see that 180cm is (almost) always within the range of each image. This can be also visualized in **Figure** *12*. Only the images DJI137, DJI43 and DJI137 don't include the real value (1.80m) within their range, assuming a GPS error of +/- 0.5m.

In other words, the accuracy in the estimation of target height depends on the error associated to the horizontal distance H. In a real case scenario, taking for example the case of the image DJI143 (first column on **Table 2**), we could only say that the real height of the target has a value included between 1.70m and 1.89m (0.19m interval).

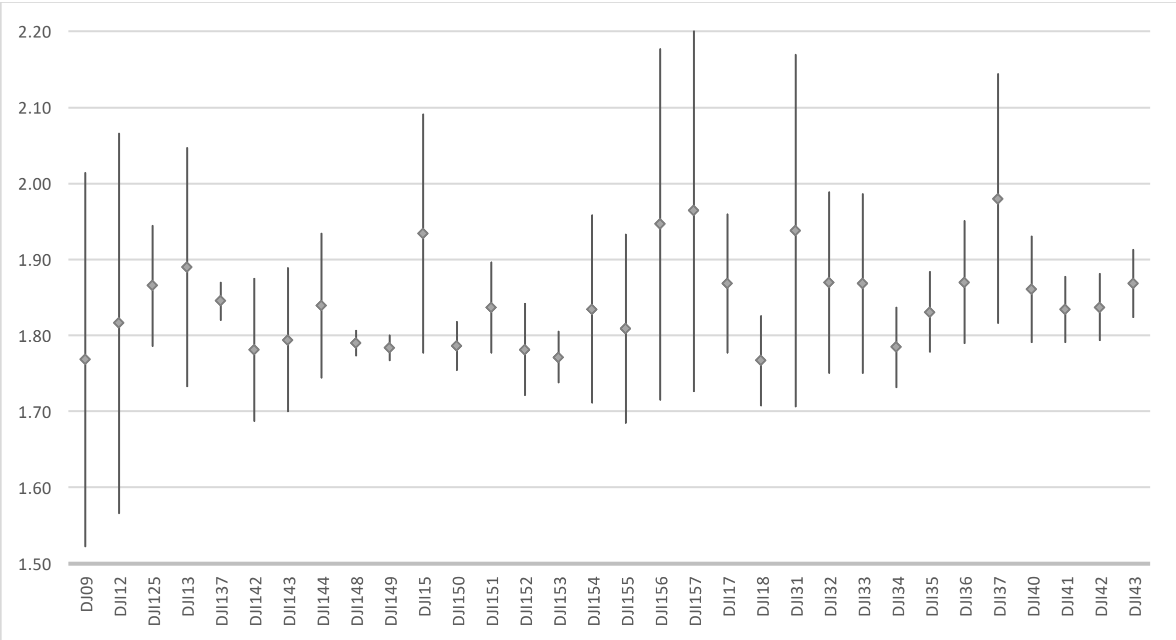


Figure 12 Grey rhombus represent the target height calculated considering GPS reading extracted from the EXIF tags for each acquired image, while the vertical lines represent the uncertainty when GPS error is 1m.

Let's now analyze the case when GPS is assumed to be 3m (the results of this analysis are not ere reported in a tabular format but only graphically in **Figure** *13*). The first element to notice is that the accuracy intervals have greatly increased, for example for DJI143 the real value may range from1.51m to 2.08m (0.57m interval), three times bigger than the interval obtained when the horizontal accuracy was assumed to be 1m. The second element to underline is that all the images have the real height value (1.80m) included in their interval. Even those images that did not have the right height within their interval assuming a positional accuracy of 1m (DJI137, DJI43 and DJI137). This simply means that the positional accuracy of those three images is more than 1m and below 3m.

In the case here under discussion the horizontal distance was calculated considering the coordinates of the target and UAV. A similar approach should be also considered when dealing with slant-range measurements obtained with laser range finders installed into UAVs. These devices may measure distances with a certain error that, taking into the account (23), generate uncertainty in the correct estimation of target height, as seen for the case above.

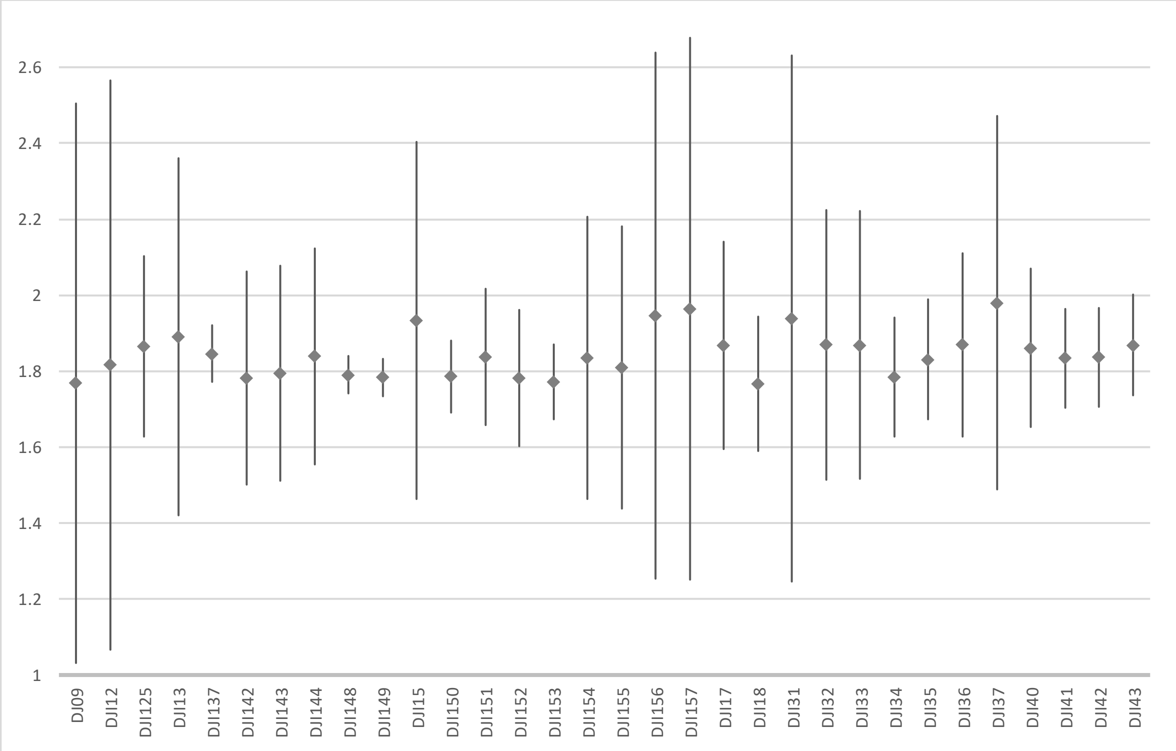


Figure 13 Grey rhombus represent the target height calculated considering GPS reading extracted from the EXIF tags for each acquired image, while the vertical lines represent the uncertainty when GPS error is 3m.

We should also notice in **Figure *12*** and **Figure *13*** that the uncertainty is not constant, but it rather changes substantially from image to image. An uncertainty analysis can be conducted using the data described in Table 1 and Table 2 to verify how the parameters involved in the calculations are affecting the uncertainty.

The first parameter to consider is the Pitch angle. However, this parameter may depend on which part of the vertical target the camera is pointing to (see **Figure *7***). To avoid this issue, it is preferable to consider the Pitch angle plus α angle, in this way we always refereeing to the bottom point of the pole in every image. In **Figure *14*** the angles obtained by the Pitch angle plus the α angle are plotted against the Height Uncertainty values of the images described in Table 1 and Table 2 (horizontal accuracy of 1m). The data distribution looks quite sparse although we may say that the uncertainty is generally growing when the Pitch angles are higher, as the best linear fit and its coefficient of determination can also attest.

The second parameter to consider is the camera-to-target horizontal distance H. If we plot in a graph the uncertainty against the horizontal distance, we can notice a clear relationship between them (see **Figure *15***). They are related by an exponential relationship which tells us that the accuracy is lower when the horizontal distance is higher. In **Figure *15*** it is reported the equation of the curve that best fits when the horizontal accuracy of 1m.

The distance from the image center to top and bottom of the pole measure in pixels after distortion correction (r_{i-corr} and r_{j-corr}) should be also considered. However, as seen for the Pitch angle, also these parameters depend on which part of the vertical target the camera is pointing to. It is therefore preferable to consider the total number of pixels spanning the feature (the pole) for this analysis. In **Figure *16*** the Total Number of Pixels for each image was plotted against the Height Uncertainty. The data distribution shows a quite evident linear trend, the equation of the best linear fit and its coefficient of determination are also reported in the figure.

Finally, intrinsic camera parameters should be also considered to analyze how they influence the overall accuracy. This analysis was not performed in this study because all the images were acquired with the same camera configuration.

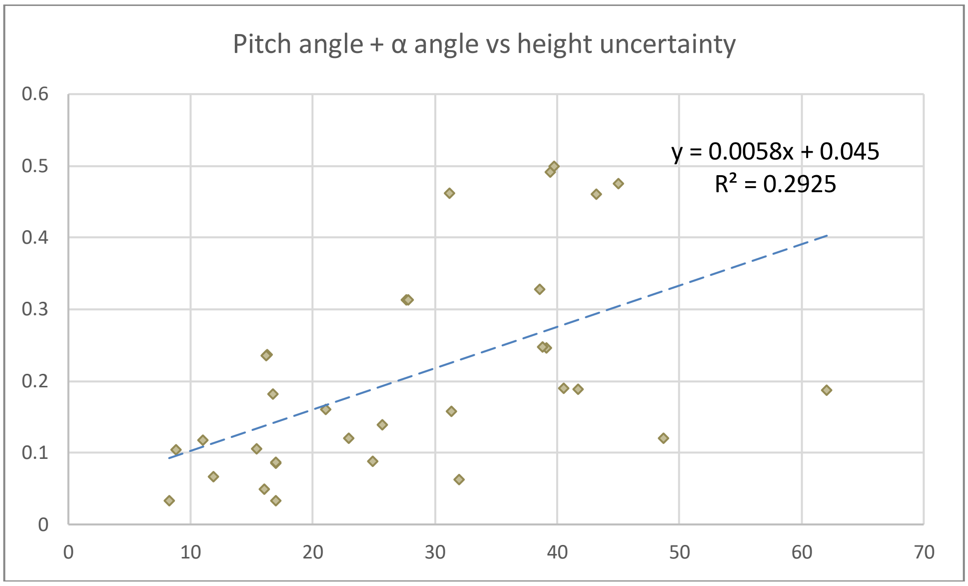


Figure 14 Pitch angle plus α angle expressed in degrees (y) vs uncertainty values expressed in meters (x) for the images described in Table 1 and Table 2. The equation of the best linear fit and its coefficient of determination are also reported.

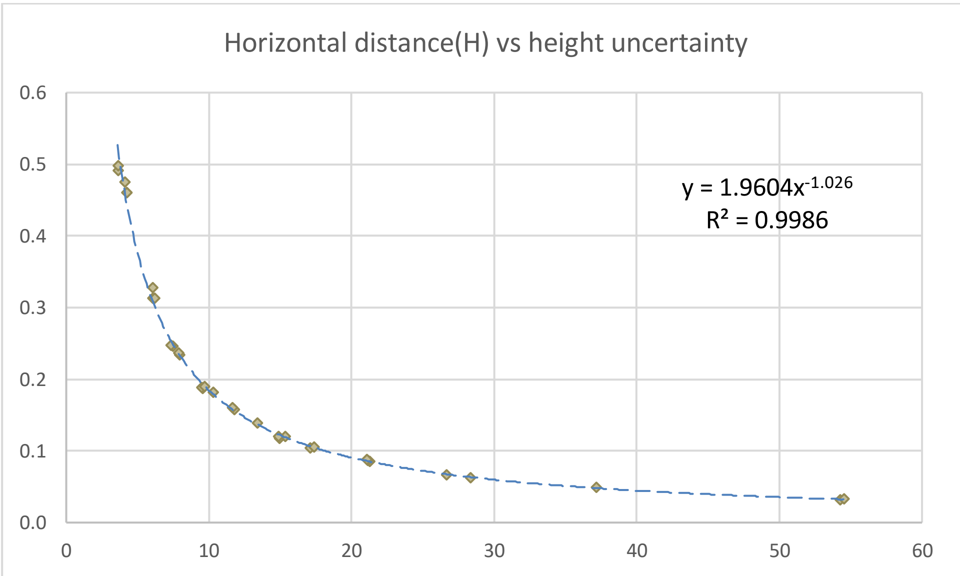


Figure 15 Horizontal distance expressed in meters (y) vs uncertainty values expressed in meters (x) for the images described in Table 1 and Table 2. The equation of the best exponential fit curve and its coefficient of determination are also reported.

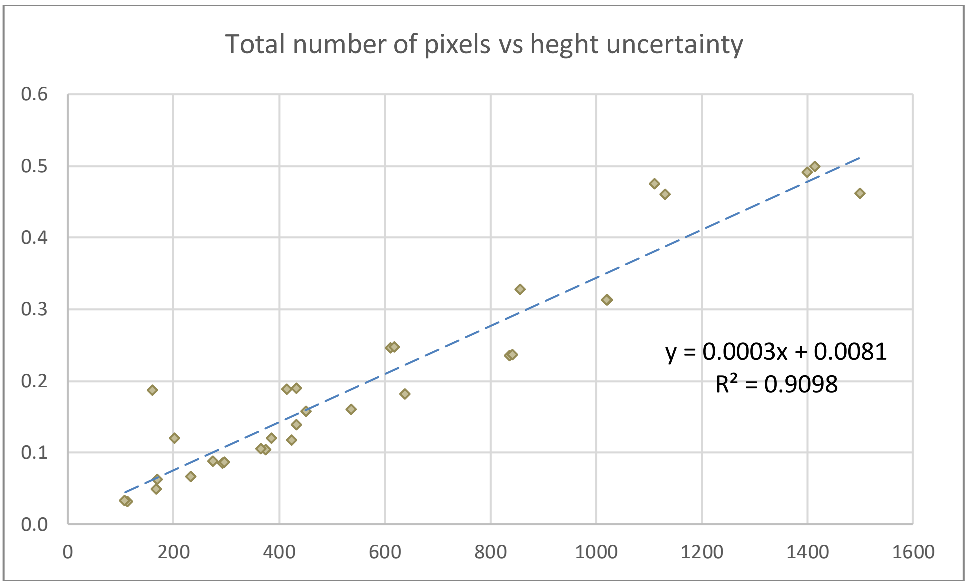


Figure 16 Total Number of Pixels (y) vs uncertainty values expressed in meters (x). The equation of the best linear fit and its coefficient of determination are also reported.

The graph in ***Figure 15*** is particularly interesting. The regression curve has an extremely good coefficient of determination to the data; therefore, the equation of the curve can tell us quite precisely which is the expected accuracy considering a certain distance from the target.

Although not directly related to uncertainty analysis, it is also interesting to analyze the correlation between Total Number of Pixels and the Horizontal Distance, this is shown in ***Figure 17***. The data distribution can be well fitted with a power regression curve, whose equation and coefficient of determination are reported in the graph. This relationship is valid for a target of 1.80m, which was the height of the pole used in this study; further research should be conducted to verify if there is a clear and well predictable relationship also for different target heights. If this is the case, the target-to-target distance could be no longer a required parameter for the calculations. The relationship shown in ***Figure 17*** also explains the linear correlation between Height Uncertainty and Total Number of Pixels seen in **Figure** *16*. On the other hand, there is no evident relationship between pith angle and number of pixels as shown in ***Figure 18***.

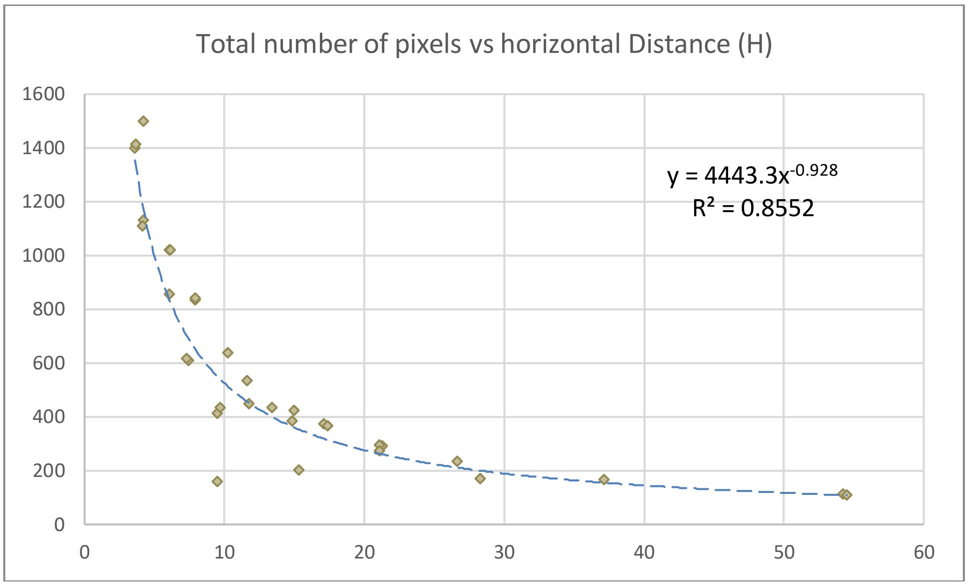


Figure 17 Total Number of Pixels (y) vs horizontal distance expressed in meters (x) for the images described in Table 1 and Table 2. The equation of the best regression curve and its coefficient of determination are also reported in the figure.

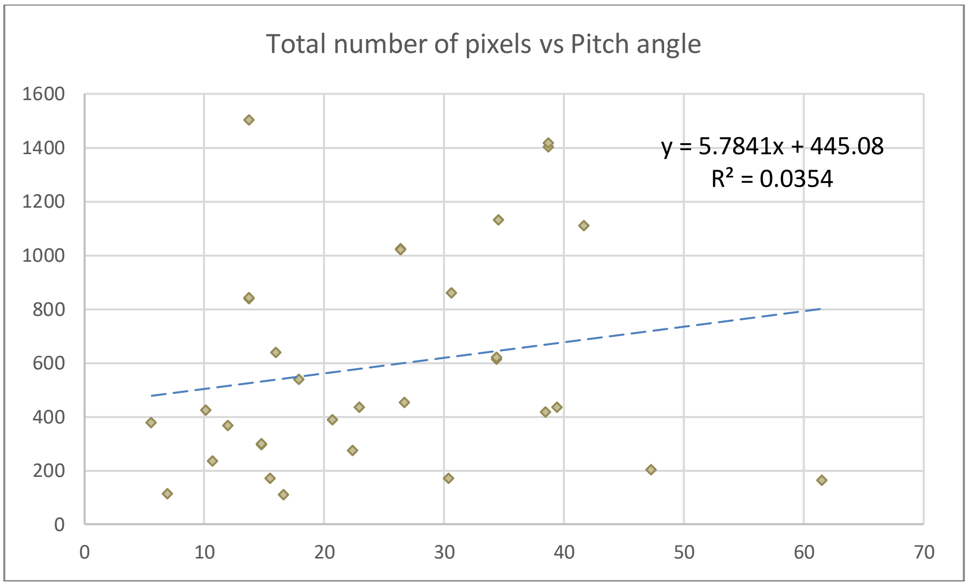


Figure 18 Total number of pixels (y) vs Pitch angle expressed in degrees (x) for the images described in Table 1 and Table 2. The equation of the best linear fit and its coefficient of determination are also reported.

Finally, it is interesting to analyze the impact of an inaccurate distortion correction. A non-accurate correction, that may lead to an erroneous value of the distance from the image center measured in pixels. Therefore, it is important to understand how a variation of number of pixels may impact the estimation of the target height. To achieve this goal, 5 pixels were added to the total number of pixels (3 upward and 2 downward, the values were chosen arbitrary) and the height was recalculated to verify the variation. The results are reported in ***Table 3***.

The difference in height varies from 0.005m up to 0.08m, and it increases linearly with the horizontal distance, as we may expect. This analysis demonstrates that a non-accurate distortion correction may lead to an erroneous value of height, which is however much lower than the error associated to the horizontal distance. This statement is anyway true for the camera and UAV system used in this study; other cameras with a different resolution may give different errors.

Table 3 The height calculated with the number of pixels measured in each image is compared with height calculated by adding 5 pixels to the original number. The table reports also the difference between the height calculated with the original number the height with when 5 pixels are added.

	Original total number of pixels	**Height (m) considering Total number of pixels**	**Total number of pixels +5**	**Height (m) considering Total number of pixels +5**	**Height Difference (m)**
DJI31	1501	1.94	1506	1.94	0.006
DJI32	843	1.87	848	1.88	0.011
DJI34	376	1.78	381	1.81	0.023
DJI41	293	1.83	298	1.86	0.031
DJ09	1401	1.77	1406	1.77	0.005
DJI12	1415	1.82	1420	1.82	0.005
DJI13	1022	1.89	1027	1.90	0.008
DJI15	1021	1.93	1026	1.94	0.009
DJI17	639	1.87	644	1.88	0.014
DJI18	424	1.77	429	1.79	0.020
DJI33	837	1.87	842	1.88	0.011
DJI35	366	1.83	371	1.86	0.024
DJI36	537	1.87	542	1.89	0.017
DJI37	858	1.98	863	1.99	0.010
DJI42	297	1.84	302	1.87	0.030
DJI125	451	1.87	456	1.88	0.019
DJI142	161	1.78	166	1.83	0.051
DJI143	415	1.79	420	1.81	0.020
DJI144	434	1.84	439	1.86	0.019
DJI148	118	1.79	123	1.86	0.074
DJI149	109	1.78	114	1.86	0.081
DJI150	170	1.79	175	1.84	0.051
DJI151	202	1.84	207	1.88	0.043
DJI152	386	1.78	391	1.80	0.022
DJI153	234	1.77	239	1.81	0.037
DJI154	611	1.83	616	1.85	0.013
DJI155	619	1.81	624	1.82	0.013
DJI156	1132	1.95	1137	1.95	0.007
DJI157	1111	1.96	1116	1.97	0.007
DJI40	434	1.86	439	1.88	0.020
DJI43	275	1.87	280	1.90	0.033
DJI137	169	1.85	174	1.90	0.054

4. Estimating the vertical length of target's subparts

In several occasions there might be the need to measure the length of subparts of the target. This is a common situation when the purpose of the surveillance activities is dedicated to estimate human body height.

In fact, the body height is the vertical distance from the bottom of bare feet to the top of the head, which shall be measured while the person is standing in a straight position [35]. If the person has a different pose, such as standing relaxed with weight on one leg or watching the telephone with the head tilted down, we would manage to estimate just the height of the body in that specific pose, not the real stature of the subject. On the other hand, in literature is well known the relationship between the height of a person its body parts (harms, hands, legs, etc.) [22] or and human face features ratio [23] obtained via experimental measures.

It is here proposed to determine pixel spacing expressed in length units (e.g. meters) measured along the vertical target, this pixel spacing can be used to estimate the length of body parts or face portions by counting the number of pixels spanning these sub-features. This is deemed to be the most practical approach to swiftly estimate the body stature when the subject does not have a straight pose in scene.

The spacing in the vertical direction is here proposed to be called Vertical Sample Distance (VSD), which can be calculated as the GSD (Ground Sample Distance, [24]), but along the vertical axis perpendicular to the ground. GSD represents the distance between pixel centers measured on the ground. When the camera is looking vertically to the ground (viewing angle τ equal to zero, see **Figure *19***), GSD can be calculated as follow [36]:

$$\mathrm{GSD} = \frac{\mathrm{d}}{\mathrm{F}}\mathrm{D} \tag{24}$$

Where d is the distance between detectors centers (pixel pitch), F is the focal length (see (3) and (4)) and D is the range. In the camera is not oriented perpendicular to the ground, τ is different than zero and the GSD must be corrected as follow, obtaining GSD′ (see **Figure 19**).

$$\mathrm{GSD}' = \frac{\mathrm{d}}{\mathrm{F}}\mathrm{D}'\frac{1}{\cos\tau} \tag{25}$$

When dealing with UAVs, the GSD′ is not a static but rather a dinamic parameter *[37]* since it changes according to the camera-to-target distance and viewing angle. Genrally speacking, we can say that each video frame has a specific GSD′.

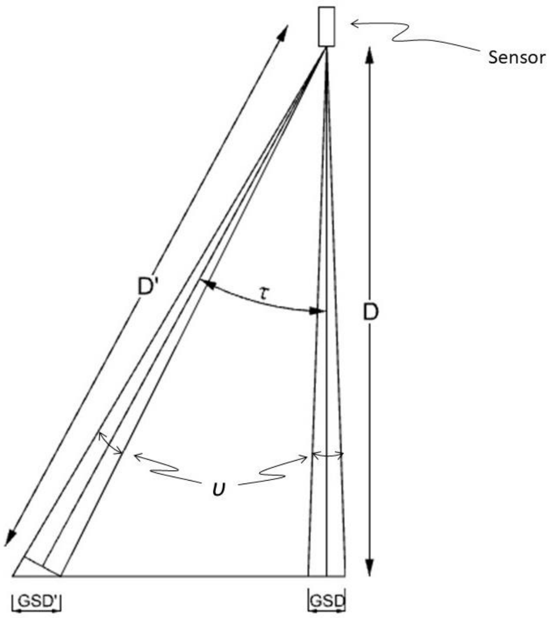


Figure 19 GSD (when the looking angle τ is zero) and GSD′ (when τ is different than zero). υ is the Field of View of a single element (pixel) of the detector.

As graphically represented in **Figure 20** and expressed in (26), the sum of all the singles GSD′ of the pixels spanning the vertical target in the image plane (distance from point I to point J) are equal to the distance p on ground (this distance has been previously mentioned during the calculations to obtain the height of the target, see (21).

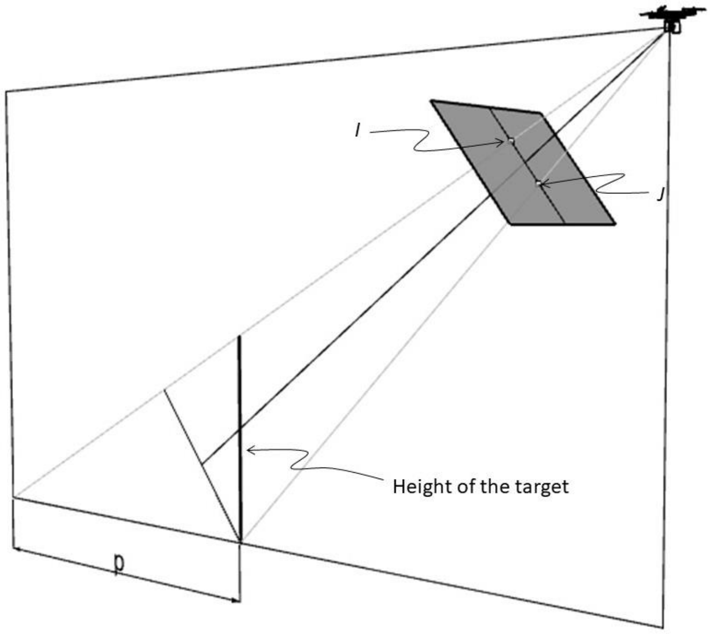


Figure 20 The horizontal distance p is the sum of the GSD′ of each pixel spanning from the point I and J in the image plane. In analogy, the height of the target.

$$\mathrm{p} = \sum_{n=i}^{j} GSD'_n \tag{26}$$

Dividing the distance p by the number of pixels spanning the feature in the image plane (I-J) would then give the average value of GSDs for each pixel within the distance I-J measured on the image place.

In this paper we propose to follow the same approach to calculate the sample distance on vertical planes, the Vertical Sample Distance (VSD), which can be calculated as follow considering **Figure 19**:

$$\mathrm{VSD}' = \frac{\mathrm{d}}{\mathrm{F}} \mathrm{D}' \sin \tau \tag{27}$$

In analogy to GSD, as graphically represented in **Figure 20** and expressed in (28), VSD of all the pixels spanning the vertical target are equal to height of the target.

$$\text{Height of the target} = \sum_{n=i}^{j} VSD'_n \tag{28}$$

Dividing the height of the target by the number of pixels spanning the feature in the image plane (I-J) would then give the average value of VSD′ for each pixel spanning the feature in the image plane (distance I-J in **Figure 20**). As seen for GSD′, also VSD′ is not constant, each image has a specific value that depends on the instantaneous camera-to-target distance and viewing angle.

More practically, we can measure the height of human subject using the procedure described in the previous chapters. If he subject does not have a straight pose, the height of the body in the specific pose can be anyway used to calculate the average VDS′ (estimated target height divided the number of pixels spanning the target) and multiply this value for the number of pixels spanning subject's subparts, such as face features or body parts, which can be used to retrieve the real stature considering ratios or relationships available in literature.

Each pixel has a specific GDS' value and, in analogy to that, each pixel has a specific VDS' value. The next paragraph intends to analyse the error generated by using the average VDS' instead the real VDS of the pixels spanning the target subpart.

4.6. Field verification

The data acquired with DJI Phantom 4 PRO can be used to verify the error due to using average VSD'. A level of 0.4m was kept tight and alighted to the pole to keep it vertical in each acquired image. The length of this level could be considered, for the sake of this analysis, as a subpart of the pole (see **Figure 11**).

For each mage (see **Table 1** and **Table 2)** the VSD was calculated dividing the real length of the pole (1.80m) by the Total Number of pixels spanning the pole. The total number of pixels spanning the level in each image was measured on screen and then corrected for lens distortion. This value was then multiplied for VSD to obtain the estimated length of the level (see **Table 4**).

Table 4 Data used to calculate the average VSD' and the estimated length of the level (real level length is 0.4m).

	Total Number of pixels spanning the pole	**Average VSD' (m) (height of the pole = 1.80m)**	**Total Number of pixels spanning the level (after distortion correction)**	**Estimated length of the Level (m)**
DJI31	1501	0.120	342	0.41
DJI32	843	0.214	191	0.41
DJI34	376	0.479	84	0.40
DJI41	293	0.614	66	0.41
DJ09	1401	0.128	320	0.41
DJI12	1415	0.127	324	0.41
DJI13	1022	0.176	230	0.41
DJI15	1021	0.176	228	0.40
DJI17	639	0.282	143	0.40
DJI18	424	0.425	93	0.39
DJI33	837	0.215	188	0.40
DJI35	366	0.492	81	0.40
DJI36	537	0.335	120	0.40
DJI37	858	0.210	194	0.41
DJI42	297	0.606	66	0.40
DJI125	451	0.399	100	0.40
DJI142	161	1.118	35	0.39
DJI143	415	0.434	91	0.39
DJI144	434	0.415	94	0.39

DJI148	118	1.525	26	0.40
DJI149	109	1.651	25	0.41
DJI150	170	1.059	38	0.40
DJI151	202	0.891	44	0.39
DJI152	386	0.466	85	0.40
DJI153	234	0.769	52	0.40
DJI154	611	0.295	133	0.39
DJI155	619	0.291	137	0.40
DJI156	1132	0.159	253	0.40
DJI157	1111	0.162	250	0.41
DJI40	434	0.415	30	0.39
DJI43	275	0.655	91	0.39
DJI137	169	1.065	37	0.39

The results in **Table 4** show that the length of the level obtained using average VDS′ gave a general error of less than 0.01m (real level length is 0.4m) which means that the average value of VDS the pixels involved in the calculation have a good average of the real value. However, this is clearly specific to the conditions analyzed in this study, namely height of the target of 1.80m, subpart of 0.4m, etc.

Finally, in a real case scenario, the average VDS′ should consider the accuracy in the estimation of the target height. Taking for example the case of the image DJI143 (first column on **Table 2**), where the estimated height of the target has an uncertainty that ranges between 1.70m and 1.89m. In this case we would get an average VSD′ included between 0.004m and 0.005m, which give a length of the level included between 0.37m and 0.41m.

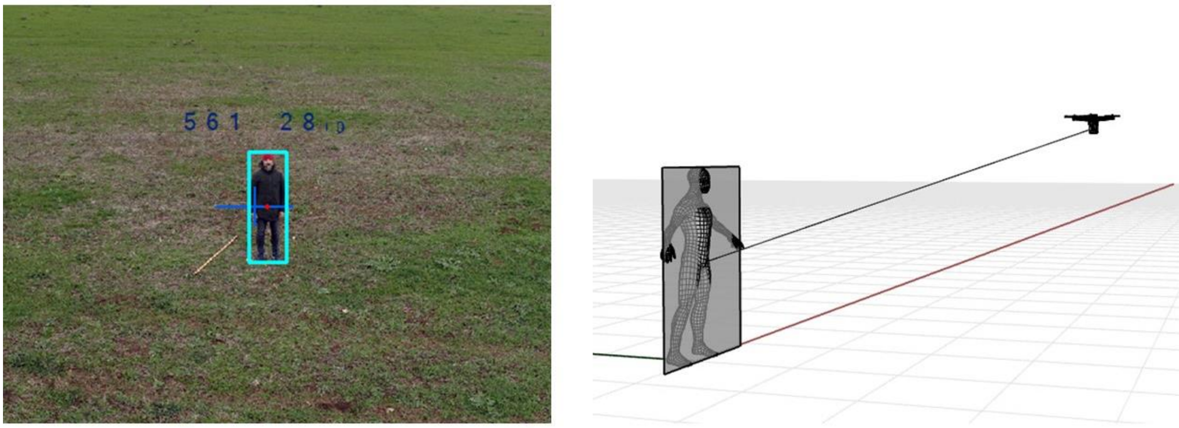
5. Discussion: Future developments for an automatic estimation of human body height in near real-time using camera installed into UAVs

Among all the possible applications of remote target height estimation using camera installed into UAVs, obtaining the real human body height is one of the most relevant. The method described in this paper, although not specifically focused on human target but usable also for other targets like trees or buildings, provides fundamental elements for this purpose.

The presence of human beings can be automatically detected in still images or video frames by computer vision algorithms, such as Histograms of Oriented Gradients (HOG) *[38]* and Viola Jones Object Detection Framework (Haar Cascades) *[39]* pre-trained for human detection. Those algorithms usually show the localization of a person via a rectangular boundary surrounding each object (see **Figure 21a**) that, if properly fitted to the human target (see **Figure 21b**), can be used to measure automatically the number of pixels spanning the target upward and downward from the image center. The camera position and orientation (camera pose) can be extracted from metadata embedded in the still images or video (e.g. KLV) to estimate in near real-time the height of the boundary box using the procedure described in this paper. If the person is standing straight in the image, the height of the bounding box could be assuming as the real stature of the subject. Otherwise, the height of the bound box divided the total number of pixels spanning the rectangle would give the average VSD′. Face identification algorithms *[40]* or other object recognition algorithms trained for body parts detection (such as arms, *[41]*) can be used to automatically detect subparts and count the number of pixels spanning vertically the face or other body parts. The number of pixels multiplied for average VDS′ would give the vertical length of those parts, which can be used to estimate the real stature of the human subject. Finally, we need always to take into the account the accuracy of the estimation to define the range within which the real height would be.

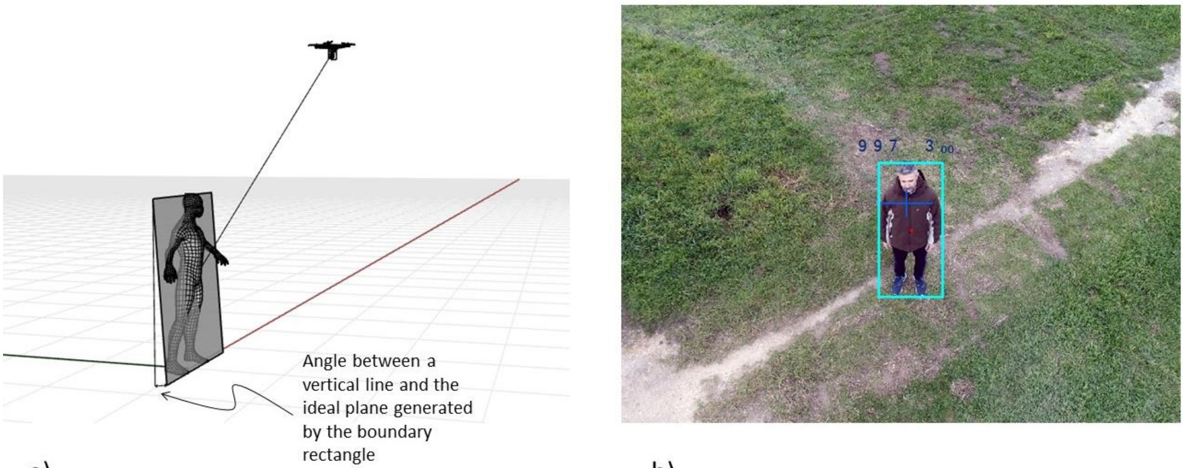
However, to achieve a solid automatic estimation of human body height in near real-time several developments still need to be implemented:

- The accuracy analysis associated to height estimation should consider also intrinsic camera parameters. In this study the intrinsic parameters were not considered because all the data (still images) were collected with the same camera.
- Algorithms previously trained for human detection available in freely available libraries may have drawbacks such as missed detection, false detections and duplicated detections for the same features. Deep Convolution Neural Networks should be used to develop more robust human detection algorithms specifically dedicated for high angle images usually acquired by UAVs *[42]*.
- Generating accurate and consistent boundary rectangles between detections of the same features in subsequent video frames is a key element to estimate real human body height. A robust detection algorithm can only partially solve this issue. In fact, high looking angles may generate non-vertical boundary rectangles (see **Figure 22**) that may introduce an additional error in estimations of human stature. Further studies are required to analysis these issues and identify suitable solutions to compensate for this problem.
- VDS can be used to estimate the length of body parts oriented vertically from the ground. However, very often, the body parts may be oriented differently (lets' think about the ordination of the legs while walking, for example). Further studies should be focused on how to measure distances in the vertical plane (see **Figure 21b)** independently from the orientation of the subpart.
- Finally, it would be very important to analyse if height can be estimated without the need of the camera-to-target distance. In fact, even a laser range finder, which is mandatory payload device when dealing with moving targets such as people, may not have enough accuracy for precise human height estimations. The graph in **Figure 17** may be of interest in this respect, since it shows a clear relationship between the total number of pixels and the horizontal distance, well descrivable by a power regression curve. This relationship is valid for a target of 1.80m, which was the height of the pole used in this study. Further studies should be conducted to analyze if there is also a clear and well predictable relatinship between the total number of pixels, horizontal distance and the height. If this is the case, the target-to-target distance could be no longer a required parameter for the calculations.



a) b)

Figure 21 a) human detection algorithms usually show the presence of person with a boundary rectangle (light blue, the red dot represents the center of the rectangle, blue cross indicates the center of the image, the number in the rectangle's top left corner indicates the total number of pixels spanning the rectangle, while the other number indicates the number of pixels from image center to the top of the rectangle). b) if the rectangle is properly fitted to the person, it can be used to measure the number of pixels spanning the target.



a) b)

Figure 22 a) Boundary rectangles generated by high looking angles may be not vertical. b) example of boundary rectangle in a real picture.

5. Conclusions

This paper described a procedure for the remote estimation of target height using gimballed camera installed into a UAV. The procedure foresees the camera calibration and image distortion compensation before using a pinhole model to calculate geometrically the vertical length of a feature. The main strengths of this approach are the simplicity and rapidness. In fact, height estimation does not require special equipment or double cameras, a single still image or video frame acquired with an optical camera is sufficient. Moreover, no vanishing lines or objects of reference height are required to be present in the scene. Finally, just few parameters are required: intrinsic camera parameters, which are usually provided by manufacturer but they can also retrieved via computer vision analysis, camera pitch angle, usually available in metadata associated to the acquired images or video data, number of pixels spanning the feature, which can be measured either manually or automatically using feature detection algorithms, and, finally, the distance between the camera and target, which can be obtained using coordinates if the position of the target is known or using devices to measure distances, such as laser range finders. These parameters are involved in simple trigonometric calculations that can be very rapidly performed for near real-time applications. Also, the processing for lens distortion compensation, which may be very time-consuming if performed for the entire image, is very swift because it involves only a very little number of pixels.

On the other hand, the weakness of this procedure is related to the uncertainty of the estimation, which is mainly linked to the error associated to the camera-to-target distance. Also, inaccurate lens distortion correction procedures may introduce some error, but this study confirmed that they are very minor compared to camera-to-sensor distance error. This distance, either acquired via coordinates difference of measured via laser range finders, is always affected by a certain error that generates an uncertainty in the height estimation. This paper has also analyzed how extrinsic camera parameters are affecting the overall uncertainty, identifying interesting relationships that may be used to define in advance the expected accuracy during surveillance activities. However, this analysis did not take into consideration how intrinsic camera parameters are concurring to the uncertainty, future studies should be dedicated to further develop this aspect. Future studies should be also dedicated on how the height can be estimated without the need of the camera-to-target distance. Along this line, this paper has shown a very well predictable relationship between the number of pixels and the camera-to-target distance, additional studies should be conducted to analyze if there is also a clear and well predictable relationship between the total number of pixels, horizontal distance and the height. If this is the case, the target-to-target distance could be no longer a required parameter for the calculations.

Another important element treated in this paper was the Vertical Sample Distance (VSD). The height a person who is not standing perfectly vertical can be derived by relationships between body parts or human face features ratio VDS can be used to estimate the length of body parts oriented vertically from the ground. However, very often, the body parts may be oriented differently (lets' think about the ordination of the legs while walking, for example). Further studies should be focused on how to measure distances in the vertical plane (see Figure 21b) independently from the orientation of the subpart.

Finally, the method described in this paper, although not specifically focused on human target but usable also for other targets like trees or buildings, provides fundamental elements for an automatic estimation of human body height in near real-time using camera installed into UAVs.

Author Contributions: Conceptualization, A.T.; methodology, A.T.; software, M.C.; formal analysis, A.T.; resources, M.P.; writing—original draft preparation, A.T.; writing—review, M.P. and M.C.; supervision, M.P. and M.C.; project administration, M.P.

Funding: This research received no external funding.

Disclaimer: The content of this paper does not necessarily reflect the official opinion of the European Maritime Safety Agency. Responsibility for the information lies entirely with the authors.

Acknowledgments: The authors would like to express their gratitude to José Cortes for the support in data acquisition, Marta Lopes for the support in data analysis, António Soares for the support in python programming and, finally, Luís Simão and Inês Simão for the logistical support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fahlstrom, P.G.; Gleason, T.J. *Introduction to UAV systems*; Aerospace series; 4th ed.; John Wiley & Sons: Chichester, West Sussex, 2012; ISBN 978-1-119-97866-4.
2. European Aviation Safety Agency Introduction of a regulatory framework for the operation of drones 2017.

3. Jung, J.; Yoo, S.; La, W.; Lee, D.; Bae, M.; Kim, H. AVSS: Airborne Video Surveillance System. *Sensors* **2018**, *18*, 1939.
4. Cai, J.; Walker, R. Height estimation from monocular image sequences using dynamic programming with explicit occlusions. *IET Computer Vision* **2010**, *4*, 149.
5. Negretti, P.; Bianconi, G.; Bartocci, S.; Terramoccia, S. Lateral Trunk Surface as a new parameter to estimate live body weight byVisual Image Analysis. *Italian Journal of Animal Science* **2007**, *6*, 1223–1225.
6. Kim, M.-G.; Moon, H.-M.; Chung, Y.; Pan, S.B. A Survey and Proposed Framework on the Soft Biometrics Technique for Human Identification in Intelligent Video Surveillance System. *Journal of Biomedicine and Biotechnology* **2012**, *2012*, 1–7.
7. Ramstrand, N.; Ramstrand, S.; Brolund, P.; Norell, K.; Bergström, P. Relative effects of posture and activity on human height estimation from surveillance footage. *Forensic Science International* **2011**, *212*, 27–31.
8. Rues, D.; Manthey, K.; Reulke, R. An accurate 3D feature tracking system with wide-baseline stereo smart cameras. In Proceedings of the 2011 Fifth ACM/IEEE International Conference on Distributed Smart Cameras; IEEE: Ghent, Belgium, 2011; pp. 1–6.
9. Calì, M.; Ambu, R. Advanced 3D Photogrammetric Surface Reconstruction of Extensive Objects by UAV Camera Image Acquisition. *Sensors* **2018**, *18*, 2815.
10. Azevedo, F.; Dias, A.; Almeida, J.; Oliveira, A.; Ferreira, A.; Santos, T.; Martins, A.; Silva, E. LiDAR-Based Real-Time Detection and Modeling of Power Lines for Unmanned Aerial Vehicles. *Sensors* **2019**, *19*, 1812.
11. Christiansen, M.; Laursen, M.; Jørgensen, R.; Skovsen, S.; Gislum, R. Designing and Testing a UAV Mapping System for Agricultural Field Surveying. *Sensors* **2017**, *17*, 2703.
12. Aguilar, W.G.; Luna, M.A.; Moya, J.F.; Abad, V.; Parra, H.; Ruiz, H. Pedestrian Detection for UAVs Using Cascade Classifiers with Meanshift. In Proceedings of the 2017 IEEE 11th International Conference on Semantic Computing (ICSC); IEEE: San Diego, CA, USA, 2017; pp. 509–514.
13. Criminisi, A.; Reid, I.; Zisserman, A. Single view metrology. In Proceedings of the Proceedings of the Seventh IEEE International Conference on Computer Vision; IEEE: Kerkyra, Greece, 1999; pp. 434–441 vol.1.
14. Hartley, R.; Zisserman, A. *Multiple view geometry in computer vision*; 2nd ed.; Cambridge University Press: Cambridge, UK ; New York, 2003; ISBN 978-0-521-54051-3.
15. Pears, N.; Liang, B.; Chen, Z. Mobile Robot Visual Navigation Using Multiple Features. *EURASIP Journal on Advances in Signal Processing* **2005**, *2005*.
16. Li, S.; Nguyen, V.H.; Ma, M.; Jin, C.-B.; Do, T.D.; Kim, H. A simplified nonlinear regression method for human height estimation in video surveillance. *EURASIP Journal on Image and Video Processing* **2015**, *2015*.
17. Jeges, E.; Kispal, I.; Hornak, Z. Measuring human height using calibrated cameras. In Proceedings of the 2008 Conference on Human System Interactions; IEEE: Krakow, Poland, 2008; pp. 755–760.
18. Fabián, J. Person Body Height Measurement by Using Surveillance Cameras. Master's Thesis, Czech Technical University in Prague: Prague, Hungary, 2015.
19. Sturm, P. Pinhole Camera Model. In *Computer Vision*; Ikeuchi, K., Ed.; Springer US: Boston, MA, 2014; pp. 610–613 ISBN 978-0-387-30771-8.
20. Zhang, Z. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2000**, *22*, 1330–1334.
21. Criminisi, A.; Zisserman, A.; Van Gool, L.J.; Bramble, S.K.; Compton, D. New approach to obtain height measurements from video.; Higgins, K., Ed.; Boston, MA, 1999; pp. 227–238.
22. Winter, D.A. *Biomechanics and motor control of human movement*; 4th ed.; Wiley: Hoboken, N.J, 2009; ISBN 978-0-470-39818-0.

23. Guan, Y.-P. Unsupervised human height estimation from a single image. *Journal of Biomedical Science and Engineering* **2009**, *02*, 425–430.
24. *Encyclopedia of optical engineering*; Driggers, R.G., Ed.; Marcel Dekker: New York, 2003; ISBN 978-0-8247-4250-8.
25. Vass, G.; Perlaki, T. Applying and removing lens distortion in post production.; 2003; pp. 9–16.
26. Basu, A.; Licardie, S. Alternative models for fish-eye lenses. *Pattern Recognition Letters* **1995**, *16*, 433–441.
27. Padhy, R.P.; Verma, S.; Ahmad, S.; Choudhury, S.K.; Sa, P.K. Deep Neural Network for Autonomous UAV Navigation in Indoor Corridor Environments. *Procedia Computer Science* **2018**, *133*, 643–650.
28. Minichino, J.; Howse, J. *Learning OpenCV 3 computer vision with Python: Unleash the power of computer vision with Python using OpenCV*; Second edition.; Packt Publishing: Birmingham, UK, 2015; ISBN 978-1-78528-384-0.
29. Motion Imagery Standards Board Motion Imagery Sensor Minimum Metadata Set 2018.
30. Redding, J.D.; McLain, T.W.; Beard, R.W.; Taylor, C.N. Vision-based target localization from a fixed-wing miniature air vehicle. In Proceedings of the 2006 American Control Conference; IEEE: Minneapolis, MN, USA, 2006; p. 6 pp.
31. Grant, B.G. *Getting started with UAV imaging systems: a radiometric guide*; SPIE Press: Bellingham, Washington, 2016; ISBN 978-1-5106-0183-3.
32. Marques, M.M. STANAG 4586: Standard Interfaces of UAV Control System (UCS) for NATO UAV Interoperability 2012.
33. DJI Phantom 4 PRO/PRO+ User Manual 2017.
34. Li, X.; Ge, M.; Dai, X.; Ren, X.; Fritsche, M.; Wickert, J.; Schuh, H. Accuracy and reliability of multi-GNSS real-time precise positioning: GPS, GLONASS, BeiDou, and Galileo. *Journal of Geodesy* **2015**, *89*, 607–635.
35. Vester, J. Estimating the Height of an Unknown Object in a 2D Image. Master Thesis, Royal Institute of Technology: Stockholm, Sweden, 2012.
36. Leachtenauer, J.C.; Driggers, R.G. *Surveillance and reconnaissance imaging systems: modeling and performance prediction*; Artech House optoelectronics library; Artech House: Boston, 2001; ISBN 978-1-58053-132-0.
37. Lazreg, N.; Bouchiha, R.; Besbes, K. Registration and correction techniques in Cubesat remote sensing images. In Proceedings of the 2017 International Conference on Engineering & MIS (ICEMIS); IEEE: Monastir, 2017; pp. 1–7.
38. Dalal, N.; Triggs, B. Histograms of Oriented Gradients for Human Detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05); IEEE: San Diego, CA, USA, 2005; Vol. 1, pp. 886–893.
39. Jia, H.-X.; Zhang, Y.-J. Fast Human Detection by Boosting Histograms of Oriented Gradients. In Proceedings of the Fourth International Conference on Image and Graphics (ICIG 2007); IEEE: Chengdu, Sichuan, China, 2007; pp. 683–688.
40. Passalis, N.; Tefas, A. Concept detection and face pose estimation using lightweight convolutional neural networks for steering drone video shooting. In Proceedings of the 2017 25th European Signal Processing Conference (EUSIPCO); IEEE: Kos, Greece, 2017; pp. 71–75.
41. Bk, S.; Corvee, E.; Bremond, F.; Thonnat, M. Person Re-identification Using Spatial Covariance Regions of Human Body Parts. In Proceedings of the 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance; IEEE: Boston, MA, USA, 2010; pp. 435–440.
42. Bejiga, M.; Zeggada, A.; Nouffidj, A.; Melgani, F. A Convolutional Neural Network Approach for Assisting Avalanche Search and Rescue Operations with UAV Imagery. *Remote Sensing* **2017**, *9*, 100.