

# HEIGHT ESTIMATION FROM A SINGLE CAMERA VIEW

Mahdi Momeni-K.<sup>1</sup>, Sotirios Ch. Diamantas<sup>2</sup>, Fabio Ruggiero<sup>3</sup> and Bruno Siciliano<sup>3</sup>

<sup>1</sup>*Ecole Centrale de Nantes, 44321 Nantes cedex 3, France.*

<sup>2</sup>*Intelligent Robot Lab, School of Electrical Engineering, Pusan National University, 609-735 Busan, Republic of Korea*

<sup>3</sup>*PRISMA Lab, Dipartimento di Informatica e Sistemistica, Università degli Studi di Napoli Federico II, Napoli, Italy*  
{mahdi\_electronic@yahoo.com, sotirios@pusan.ac.kr, {fabio.ruggiero, bruno.siciliano}@unina.it}

Keywords: Height Estimation, Monocular Camera, Vanishing Point

Abstract: This paper presents a novel technique for the estimation of the height of an object using a single camera view. In the proposed method, the only information required is the knowledge about the pose of the camera with respect to the world (i.e., height and pitch angle of the camera with respect to the ground) and a vanishing point. In the developed theory, the focal length may also be known, but in the proposed experiments it has not been employed: an approximation for small pitch angles has been taken into account and the consequent committed error has been then analysed. The presented method gives accurate results for any object placed in unstructured environments, regardless of the relative distance from the camera. The method has been tested in a series of outdoor and indoor environments, and the experimental results are presented in this paper.

## 1 INTRODUCTION

Visual information constitutes a core part in image processing field and in robotics science. The application areas are numerous and range from single view metrology, object recognition, autonomous robot navigation, visual servoing and so on.

In this paper, the problem of height estimation is addressed. In particular, the height of an object, irrespective of its distance from the camera, is computed by using only the knowledge about the pose of the camera (i.e., height and pitch angle of the camera with respect to the ground) and a vanishing point, no matter if this last is in the image plane or not. Focal length should also be known in theory, but in practical application it is employed when large pitch angles arise. Moreover, the object can be placed in any kind of environment since the proposed method can easily employ some rough object detection techniques in order to recognize the object in the scene – provided that the object is distinguishable from the background. Hence, in the proposed methodology, it is not required to estimate the vertical vanishing lines, the reference direction vanishing point, and the 3D properties from the 2D images are not detected.

The employed assumptions are very often fulfilled in the aforementioned applications. For instance, in robotics, a UAV usually employs an attached camera in order to retrieve some information about the exter-

nal environment (Nonami et al., 2010). Hence, for instance, the proposed method could be useful in order to detect the objects' height during the vehicle's path so as to avoid them during the flight. The position of the camera can be recovered by knowing the vehicle attitude and altitude with respect to the ground. Other important applications can be considered the image processing field, robot navigation (Angelova et al., 2006), single view metrology.

This paper comprises five sections. Following is Section 2 where the related works in height estimation and their application areas are presented. In Section 3 the methodology of estimating object height is described. Section 4 presents the results of the approach and gives a comparison with existing methods. Finally, Section 5 epitomizes the paper with a discussion on the conclusions drawn from this research as well as the future prospects of this method.

## 2 RELATED WORKS

A significant number of works tackles the problem of height estimation using calibrated cameras. (Viswanath et al., 2009) describe a simple model of error distribution in the computed height as a function of the location of the object and the recovered height of the camera. In their approach they estimate the vanishing points and the horizon line in order to

compute the height of the camera. Upon completion of this task, they calculate the height of the object of interest and then they select an error model in order to correct the estimated height. Their approach is demonstrated on an object tracking scenario. (Criminisi et al., 2000) describe various methods for estimating the height of an object using an uncalibrated camera. In their work, they describe how the affine 3D geometry of a scene can be measured by a single perspective image. Their approach is based on the estimation of the vanishing line of the ground plane and the vertical vanishing point. Nevertheless, errors appear in the estimated height of the object since these accrue from errors in estimating the ground plane and in estimating the 3D position of the object. In a more recent attempt, (Pears et al., 2005) propose a method for height estimation using the height of an object as a reference in order to correctly scale affine height measurements to Euclidean ones. In their approach, the authors segment the ground plane from the visual field of view of a mobile robot and estimate the height of objects that lie above the ground plane of the robot. In addition to the previous work, (Chen et al., 2006) present a visual metrology method by using uncalibrated images separated by a near pure translation. This approach has been applied to two cases, one for detecting obstacles in a mobile robot application and a second one for measuring the height of doorways, windows, and other furniture. In the first case the authors use the camera height as a known parameter whereas in the second case they use a reference object with a known height. (Cai and Walker, 2010) estimate the height of trees and power lines from an unmanned aerial vehicle. In this research sequential images are captured by a single camera and with known distances of at least two objects are able to infer the height of the objects. The use of dynamic programming for stereo matching provides depth information in occluded regions.

Human height estimation has been a thriving field of interest the last few years with applications in forensic image analysis, video surveillance, and human tracking. (BenAbdelkader and Yacoub, 2008) present a method where they are able to estimate the height of human body without using a reference length in the image plane. In addition, their estimation can be achieved when the upper body of a human is visible. For their method they use a statistical method in the Bayesian-like framework. The intrinsic and extrinsic parameters of the camera need not to be known, and thus the camera may even be uncalibrated. In a similar case, (Jeges et al., 2008) estimate the human height using a calibrated camera. In this research the authors estimate the height

of humans along with the process of calibrating the camera. Their application area has been in tracking humans in intelligent surveillance systems. (Guan, 2009) presents a method for estimating the human stature based on the human face. In this approach a single image and a calibrated camera are used for estimating the stature. The author takes into consideration the ratios of features in the human face to estimate the stature of a human: in particular, the vertical proportions of the human face remain unchanged during human growth. Thus, based on the golden section rule that governs the human face, an estimation of the human stature is achieved.

In robot navigation the extraction of vertical lines has been used by (Zhou and Li, 2007) where the authors exploit the vertical lines of an indoor environment in order to make the robot able to navigate autonomously. Furthermore, based on the camera orientation which is inferred by the extraction of vertical lines, their system detects and hence avoids obstacles. Finally, (Sun et al., 2006) make use of stereo vision in order to infer the height of vegetation from the powerlines. The proposed method makes use of images captured by an airborne system and a stereo vision system that recovers the surface of vegetation. The system identifies power poles and measures the distance between the vegetation and the modelled line.

### 3 METHODOLOGY

This section presents the method followed for the estimation of the height of an object. First, the notation and the basic assumptions are presented, then the key theorems are introduced, while in the remainder the core of the methodology is described.

#### 3.1 Basic assumptions

The premise to find vanishing points is to consider environments which contain parallel straight lines. When looking at the perspective projection of 3D parallel lines, they intersect in one common point in the image plane, which is the so called vanishing point (Hartley and Zisserman, 2004). Vanishing points, therefore, represent the projections of 3D points laying at infinity, since parallel lines intersect at infinity (Saurer et al., 2010). However, depending on the orientation of the camera, the vanishing point may be either inside or outside of the camera field of view. Further, a camera can be characterized by the so called intrinsic and extrinsic parameters (Siciliano et al., 2008): the former depend on the lens and sensor characteristics; the latter on the relative pose between

a frame attached to the camera and a fixed reference frame, usually attached to the ground.

The goal of this work is to find the height of an object in the observed scene on the basis of the following assumptions:

- It is possible to measure both the height of the camera and the angle of the optical axis (i.e. *pitch angle*) with respect to the ground.
- It is possible to extract a vanishing point.
- In the developed theoretical framework, the camera focal length should be known. This information can be retrieved from the camera datasheet, since this is not the usual  $(2 \times 1)$  vector that comes out from the common calibration procedures. However, in the experiments, this assumption will be relaxed.
- If automatic techniques about object recognition would be used in the experiments, then the object has to be distinguishable from the background.

In order to compute the height of an object in meters, its height in pixel has to be known. Then, the key of the methodology is to obtain the ratio between pixel and meters in the observed scene: by knowing the camera height in the real 3D world, it would be possible to compute such a ratio if it is allowed to recognize the camera height in pixels in the image plane. Since the further away the object from the camera is, the smaller the size it has on the image plane, then the pixel-height of the camera should be computed at the same position of the object in the image plane.

### 3.2 Basic theorems

With reference to Fig. 1, consider a world fixed frame  $\Sigma_w = O_w - x_w y_w z_w$  and a frame  $\Sigma_c = O_c - x_c y_c z_c$  attached to the camera, whose origin  $O_c$  is at the lens center; further, consider a reference frame on the image plane, whose axes  $U$  and  $V$  are parallel to the axes  $x_c$  and  $y_c$  of  $\Sigma_c$ , and whose origin is at the intersection of the optical axis with the image plane, termed principal point. Let  $\beta$  be the pitch angle with respect to the ground, that is the angle between the axes  $z_c$  and  $z_w$  around  $x_w$ , and where positive rotations are considered counter-clockwise. Finally, consider a fixed virtual frame  $\Sigma_v = O_v - x_v y_v z_v$  that is coincident to the camera frame when the pitch angle  $\beta$  is zero. It comes out that the pitch angle can also be seen as the angle between the axes  $z_c$  and  $z_v$  around  $x_v$ .

By considering the frontal perspective projection model – radial distortion can be neglected since this does not affects the method (Guan, 2009) –, a point  $p^c = [p_x^c \ p_y^c \ p_z^c]^T$ , whose coordinates are expressed with respect to  $\Sigma_c$ , will project from

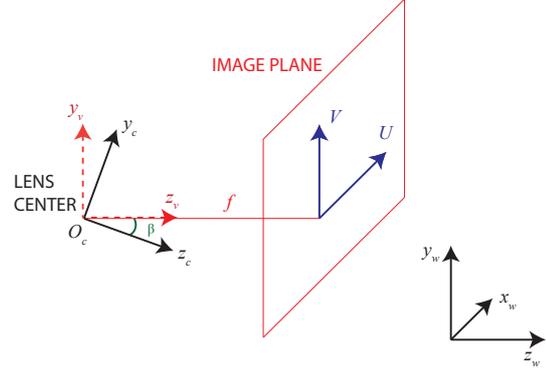


Figure 1: Significant frames employed in the proposed methodology. In order to simplify the figure, the axes  $x_c$  and  $x_w$  have not been drawn, while the image plane has been depicted in the condition  $\beta = 0$ .

the 3D scene onto the image plane with coordinates  $[u \ v]^T$  in the  $UV$ -frame through the following expression (Hutchinson et al., 1996):

$$[u \ v]^T = (f/p_z^c) [p_x^c \ p_y^c]^T \quad (1)$$

where,  $f$  is the focal length of the camera. All objects in the field of view of the camera, therefore, will project onto the 2D image plane through the nonlinear equation in (1).

Suppose now to have two points with equal component on the axes  $x_c$  and  $z_c$  in the camera frame, but different components on the  $y_c$  axis: the following theorem can be thus introduced.

**Theorem 1.** *Although two points are projected from the 3D scene onto the 2D image plane through the nonlinear equation (1), the ratio between the two different  $y_c$  components will remain unchanged, as long as they have the same values for the other two components, with all the quantities expressed with respect to the camera frame.*

*Proof.* Two points  $P_1 = [p_{x_1}^c \ p_{y_1}^c \ p_{z_1}^c]^T$  and  $P_2 = [p_{x_2}^c \ p_{y_2}^c \ p_{z_2}^c]^T$  are projected onto the image plane through (1). Since the two points have the property that  $p_{z_1} = p_{z_2}$  and  $p_{x_1} = p_{x_2}$ , it is possible to write

$$\begin{cases} u_1 = (f/p_{z_1})p_{x_1} \\ v_1 = (f/p_{z_1})p_{y_1} \end{cases} \quad \begin{cases} u_2 = (f/p_{z_1})p_{x_1} \\ v_2 = (f/p_{z_1})p_{y_2} \end{cases} \quad (2)$$

By comparing the two systems in (2), it yields to

$$v_1/v_2 = p_{y_1}/p_{y_2}, \quad (3)$$

with  $u_1 = u_2$ , and this proves the theorem.  $\square$

The direct result of Theorem 1 is that the ratio of the  $y_c$ -coordinates in the camera frame of any two

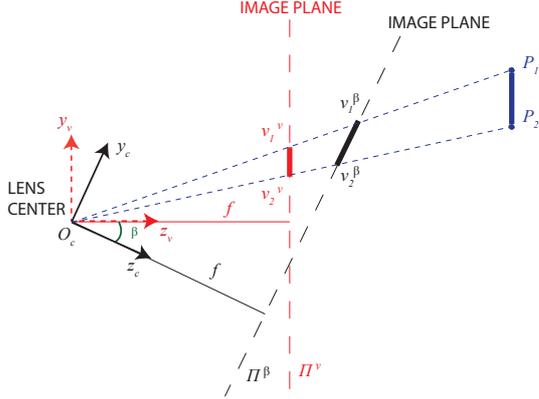


Figure 2: Frontal perspective projection of a segment in two images planes with a different camera orientation. In red the case of  $\beta = 0$  (i.e.,  $\Sigma_c$  is coincident with  $\Sigma_v$ ), while in black the case of a generic pitch angle  $\beta$ .

points laying along a line parallel to the axis  $y_c$  in the 3D world, remains equal to the ratio of the  $v$ -coordinates once these same points are projected in the 2D image plane.

Now, it is introduced a theorem that relates the projections of a point from the 3D world onto two different image planes.

**Theorem 2.** *By knowing the  $v$ -coordinate  $v^\beta$  of the projection of a 3D point  $P$  onto an image plane  $\Pi^\beta$ , associated with a camera whose optical axis has an angle  $\beta$  with respect to ground, it is possible to compute the  $v$ -coordinate  $v^v$  of the projection of the same 3D point onto the image plane  $\Pi^v$ , associated with a virtual camera whose frame is coincident with the  $\Sigma_v$ . In particular, the following relationship holds:*

$$v^v = f(v^\beta \cos \beta - f \sin \beta) / (v^\beta \sin \beta + f \cos \beta). \quad (4)$$

*Proof.* With reference to Fig. 2, consider a point  $P$  in the 3D space, the virtual reference frame  $\Sigma_v$  and a camera frame  $\Sigma_c$  whose optical axis has a pitch angle  $\beta$  with respect to the ground. The  $v$ -coordinate of the projection of the point  $P$  onto the image plane  $\Pi^\beta$  associated to  $\Sigma_c$  is denoted with  $v^\beta$ . Without loss of generality, consider that the  $u$ -coordinate of the projection of the point  $P$  is zero.

From the definition of the frontal perspective model, the 3D coordinates of the projection of the point  $P$  onto  $\Pi^\beta$  with respect to the camera frame  $\Sigma_c$  is represented by the vector  $p_\beta^c = [U \ V \ f]^T = [0 \ v^\beta \ f]^T$ . Pre-multiplying such a vector for the well-known expression of the rotation matrix  $R_c^v$  (Siciliano et al., 2008) that represents the rotation of the frame  $\Sigma_c$  with respect to the frame  $\Sigma_v$  around the axis  $x_v$ , it is thus possible to express  $p_\beta^c$  with respect to the frame  $\Sigma_v$ . Hence,  $p_\beta^v = R_c^v p_\beta^c = [0 \ v^\beta \cos \beta - f \sin \beta \ v^\beta \sin \beta + f \cos \beta]^T$ .

The optical ray that passes through the lens center, the point  $P$  and both the projections of  $P$  onto  $\Pi^\beta$  and  $\Pi^v$  is then represented by the following straight line equations with respect to the frame  $\Sigma_v$

$$\begin{cases} x_v = 0 \\ y_v = (v^\beta \cos \beta - f \sin \beta) t \\ z_v = (v^\beta \sin \beta + f \cos \beta) t \end{cases}, \quad (5)$$

where  $t$  parametrizes such straight line. The interception of (5) with the image plane  $\Pi^v$ , whose equation is  $z_v = f$  with respect to the frame  $\Sigma_v$ , yields to

$$f = (v^\beta \sin \beta + f \cos \beta) t \quad (6)$$

that gives the projection of the point  $P$  onto  $\Pi^v$ . By computing  $t$  from (6) and by substituting it into  $y_v$  in (5), it is hence possible to prove the theorem.  $\square$

From the above Theorem 2 it is possible to retrieve the following corollary.

**Corollary 1.** *With reference to Fig. 2, by knowing the  $v$ -coordinates  $v_1^\beta$  and  $v_2^\beta$  of the projections of two 3D points  $P_1$  and  $P_2$ , respectively, onto the image plane  $\Pi^\beta$ , it is possible to compute the  $v$ -coordinates  $v_1^v$  and  $v_2^v$  of the same two 3D points  $P_1$  and  $P_2$ , respectively, onto the image plane  $\Pi^v$ , that is as the camera had no pitch angle with respect to the ground. In particular such relation holds*

$$|v_1^v - v_2^v| = f^2 |v_1^\beta - v_2^\beta| / \gamma_\beta, \quad (7)$$

where  $\gamma_\beta = |(v_1^\beta \sin \beta + f \cos \beta)(v_2^\beta \sin \beta + f \cos \beta)|$ , and where  $|\cdot|$  denotes the absolute value operation.

The direct result of the Theorem 2 and of the Corollary 1 is that if it is possible to compute the length of a segment in pixel in an image plane whose camera frame has a certain pitch angle  $\beta$ , it is then possible to have the pixel length of the same segment as the camera had the optical axis parallel to the ground.

### 3.3 Methodology steps

Another direct consequence of the Theorem 1 is that the ratio of the height of any two adjacent objects in image plane represents the ratio of the two adjacent objects in the real 3D environment. As a result, by knowing the proportion of camera height w.r.t. the object height, at the object's position in the image plane, it would be possible to estimate the height of the object in the real world. To this end, denoting with  $H_c$  the height of the camera with respect to  $\Sigma_w$ ,  $H_{c_p}$  the height of the camera in pixel,  $H_o$  the height of the object with respect to  $\Sigma_w$  and  $H_{o_p}$  the height of the object in pixel, the following relationship holds

$$H_c / H_{c_p} = H_o / H_{o_p}. \quad (8)$$

The information about  $H_c$  is given by the assumptions, while  $H_{op}$  can be retrieved by the acquired picture, thus the problem is to compute the height of the camera  $H_{c_p}$  in the image plane.

Consider the frame  $\Sigma_v$ , the associated image plane  $\Pi^v$  and a plane  $\Pi^g$  parallel to the ground that contains also the axis  $z_v$ : this plane and the ground plane meet each other at infinity and they form a line that contains a vanishing point. Since the optical lens center, that is the origin of  $\Sigma_v$ , belongs to  $\Pi^g$ , this plane and thus the vanishing point can be used as a reference in order to find out the height of the camera in the image plane  $\Pi^v$ . In particular, this implies the following Lemma.

**Lemma 1.** *The distance between a point on the ground plane and the  $v$ -coordinate of the vanishing point in the image plane, and thus the horizontal line that passes through the vanishing point, represents the height of the camera in pixel.*

Hence, the above Lemma 1 implies that, in the image plane  $\Pi^v$ , the  $v$ -distance between the ground at the object position and the vanishing point represents the height of the camera in pixel. By computing also the height of the object in pixel, it would be possible to apply equation (8), thanks to Theorem 1, in order to get the height of the object in the 3D world. However, the above Lemma 1 has been derived with respect to  $\Sigma_v$ , while in a general case the camera can have a pitch angle with respect to the ground. Hence, it is only possible to compute in pixel both the camera and the object height into the image plane  $\Pi^\beta$ . In order to follow Lemma 1 and thus to apply equation (8), it is necessary to employ equation (7) so as to refer to all the quantities with respect to the virtual frame  $\Sigma_v$ .

Then, to recap, the following steps are proposed in order to get the height of an object:

1. Take a picture of the current scene with the object in the camera field of view. Then, extract the vanishing point and the outlines of the object in the image plane  $\Pi^\beta$ .
2. The  $v$ -distance between the lower end-point and the upper end-point of the object's outline represents the height of the object in pixel.
3. The  $v$ -distance between the lower end-point's corresponding point on the ground plane and the  $v$ -coordinate of the vanishing point represents the height of the camera in pixel at object's position.
4. Use equation (7) in order to refer to the above quantities with respect to  $\Sigma_v$ , so as to project the segments from  $\Pi^\beta$  into  $\Pi^v$ , in order thus to get  $H_{op}$  and  $H_{c_p}$ , respectively.
5. Use equation (8) in order to estimate the height of the object  $H_o$ .

## 4 EXPERIMENTS

This section presents the results about the proposed method, based on experimental work in both indoor and outdoor environments.

### 4.1 Technical details

The experimental set-up implementing the proposed methodology for the indoor environment is composed of an industrial USB iDS UEYE UI-1220SE-C camera mounted directly on the end-effector of a Comau Smart-Six robot manipulator, that it has been employed in order to carry around the camera in all the acquisition stations. By using such a robotic set-up, it is easier to retrieve the information about the pitch angle  $\beta$  and the height of the camera  $H_c$ . The acquired images have a resolution of  $(752 \times 480)$  pixel.

For what concerns the outdoor experiment, instead, the pocket camera SONY DSC-TX9 with 12.2 megapixels resolution has been employed, while a tripod with spirit level has been used so as to verify that the camera was parallel to the ground.

The outlines of the object have been extracted manually, since the employed objects were not so clearly distinguishable from the background. In particular, the segmentation algorithm used in (Hoiem et al., 2007), has been tested in order to automatically detect the objects' contours, but the obtained results were not so robust.

The vanishing point has also been extracted manually from the captured images. However, with the automatic methods explained by (Kongand et al., 2009), (Rother, 2002), (Schmitt and Priebe, 2009), it has been verified that the obtained results were robust enough even by using such algorithms.

It is worth noting that equation (7) can be simplified if small pitch angles are employed:

$$|v_1^v - v_2^v| \simeq |v_1^\beta - v_2^\beta|, \quad \text{when } \beta \simeq 0, \quad (9)$$

thus, in those cases, it is useless to consider the focal length  $f$ . Hence, the following experiments have been performed by considering small pitch angles, so as to stress the proposed methodology. The error in pixel that is carried out is as follows

$$e = f^2 |v_1^\beta - v_2^\beta| / \gamma_\beta - |v_1^\beta - v_2^\beta|, \quad (10)$$

obtained by subtracting the employed approximation (9) from the theoretical value (7).

### 4.2 Results and discussion

In Table 1 the results obtained during the experiments have been summarized: the first column denotes the

Table 1: Table of results

	<i>Cam.</i>	<i>Env.</i>	<i>Cam.'s H.</i>	$\beta$	<i>Real H.</i>	<i>Estim. H.</i>	<i>Error</i>
Fig. 3.1	USB	I	45.15 cm	0°	36 cm	36.3 cm	0.8%
Fig. 3.2	USB	I	107.25 cm	-7°	29.6 cm	29.9 cm	1%
Fig. 3.3	USB	I	79.87 cm	-10°	36 cm	35.7 cm	0.8%
Fig. 3.4	USB	I	79.87 cm	-10°	8.1 cm	8 cm	1.2%
Fig. 3.5	USB	I	168.90 cm	-15°	29.6 cm	29.1 cm	1.7%
Fig. 3.6	USB	I	134.42 cm	-20°	12.3 cm	11.4 cm	7.3%
Fig. 4	TX9	O	115 cm	0°	176 cm	175.6 cm	0.2%

considered figure; in the *Cam.* column it is reported the employed camera model for the current picture; the symbols I or O under column *Env.* denote an indoor or outdoor environment respectively, while the height of the camera is mentioned in the *Cam.'s H.* column; the pitch angle is reported in the column  $\beta$ ; the real height of the object and the estimated height are given in the *Real H.* and *Estim. H.* columns, respectively, and finally, in the *Error* column, the percentage of error with respect to the real height of the considered object is underlined.

In Figs. 3.1 and 4 the vanishing point is inside the image plane, while in Figs. 3.2-3.6 it is outside the image plane. A small change of the illumination can be noticed throughout all the figures, and this does not affect the method, as long as the object contours extraction methods are robust enough. In Figs. 3.2 and 3.5, the object is located on top of a cupboard so that three objects can be considered: the box, the cupboard and the “total” object given by the box over the cupboard. In order to estimate the height of the box, first the total height of the box over the cupboard, then the height of the cupboard have to be estimated: thus, by deducting the latter from the former, it is possible to compute the height of the box. Since the heights of two objects have to be computed to have in turn the desired height, because of the small pitch angles assumption, the total error on the final result is the sum of the errors after the two previous estimations. Further, it can be noticed the larger is  $\beta$ , the larger is the error: this comes out by considering (9) in lieu of (7). By considering the focal length  $f$ , it would be possible to reduce such errors.

On the basis of the previous experimental results, it can be seen that the method is general and yields to results which are achieved without the need of any other reference inside and/or outside of the image plane. The method by (Criminisi et al., 2000) yields to accurate results too, but a reference height in the image plane is used. As an example, in order to estimate the height of a person who is standing beside the phone box (refer to Fig. 10 in (Criminisi et al., 2000)), not only they need to detect the vanishing line of the ground plane, but also they need to detect the edges of the phone box in order to compute the vertical van-

ishing point. Whereas, using the proposed method, this goal can be achieved with only the previously declared assumptions (see the experiment performed in Fig. 4). On the other hand, (Guan, 2009) presents a method for human height estimation based on the human face. In his approach, the author exploits the fact that the vertical proportions of the human face remain unchanged during human growth. He also used a calibrated camera in order to achieve this task. Although his method is accurate enough, it is dependent on accurate camera calibration: besides this, a face recognition program is needed.

## 5 CONCLUSION AND FUTURE WORK

In this paper, an efficient and robust method for estimating the object height using the camera pose with respect to the world and a vanishing point has been described. The accuracy of the proposed method has been experimentally proven in outdoor and indoor environments. Future work will be focused on taking into account also roll and yaw rotations of the camera, and by considering a most automated process.



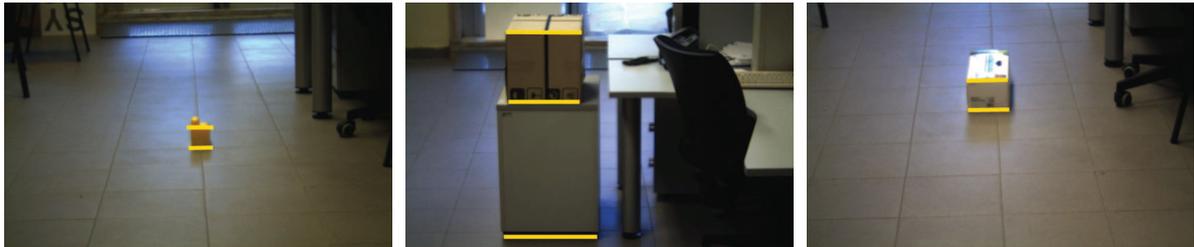
Figure 4: Actual height of the person is 176 cm, estimated height is 175.6 cm, and  $\beta = 0^\circ$ .

## REFERENCES

- Angelova, A., Matthies, L., Helmick, D., and Perona, P. (2006). Slip prediction using visual information.



3.1: Actual height of the object is 36 cm, 3.2: Actual height is 29.6 cm, estimated height is 36.3 cm, and  $\beta = 0^\circ$ . 3.3: Actual height is 36 cm, estimated height is 29.9 cm, and  $\beta = -7^\circ$ . 3.3: Actual height is 36 cm, estimated height is 35.7 cm, and  $\beta = -10^\circ$ .



3.4: Actual height is 8.1 cm, estimated height is 8 cm, and  $\beta = -10^\circ$ . 3.5: Actual height is 29.6 cm, estimated height is 29.1 cm, and  $\beta = -15^\circ$ . 3.6: Actual height is 12.3 cm, estimated height is 11.4 cm, and  $\beta = -20^\circ$ .

Figure 3: Collection of the pictures about the indoor experiments.

In *Proceedings of Robotics: Science and Systems*, Philadelphia, USA.

- BenAbdelkader, C. and Yacoob, Y. (2008). Statistical body height estimation from a single image. In *Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 1–7.
- Cai, J. and Walker, R. (2010). Height estimation from monocular image sequences using dynamic programming with explicit occlusions. *IET Computer Vision*, 4(3):149–161.
- Chen, Z., Pears, N., and Liang, B. (2006). A method of visual metrology from uncalibrated images. *Pattern Recognition Letters*, 27(13):1447–1456.
- Criminisi, A., Reid, I., and Zisserman, A. (2000). Single view metrology. *International Journal of Computer Vision*, 40(2):123–148.
- Guan, Y.-P. (2009). Unsupervised human height estimation from a single image. *Journal of Biomedical Science and Engineering*, 2(6):425–430.
- Hartley, R. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition.
- Hoiem, D., Efros, A., and Hebert, M. (2007). Recovering surface layout from an image. *International Journal of Computer Vision*, 75(1):151–172.
- Hutchinson, S., Hager, G. D., and Corke, P. I. (1996). A tutorial on visual servo control. *IEEE Transactions on Robotics and Automation*, 12(5):651–670.
- Jeges, E., Kispal, I., and Hornak, Z. (2008). Measuring human height using calibrated cameras. In *Proceedings of the 2008 Conference on Human Systems Interactions*, pages 755–760.
- Kongand, H., Audibert, J., and Ponce, J. (2009). Vanishing point detection for road detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009., pages 96–103.
- Nonami, K., Kendoul, F., Suzuki, S., Wang, W., and Nakazawa, D. (2010). *Autonomous Flying Robots*. Springer, first edition.
- Pears, N., Liang, B., and Chen, Z. (2005). Mobile robot visual navigation using multiple features. *Journal on Applied Signal Processing*, 14:2250–2259.
- Rother, C. (2002). A new approach to vanishing point detection in architectural environments. *Image and Vision Computing*, 20(9-10):647–655.
- Saurer, O., Fraundorfer, F., and Pollefeys, M. (2010). Visual localization using global visual features and vanishing points. In *Proceedings of the Conference on Multilingual and Multimodal Information Access Evaluation*, pages 1–9.
- Schmitt, F. and Priese, L. (2009). Vanishing point detection with an intersection point neighborhood. In *Discrete Geometry for Computer Imagery*, pages 132–143.
- Siciliano, B., Sciavicco, L., Villani, L., and Oriolo, G. (2008). *Robotics. Modelling, Planning and Control*. Springer, London.
- Sun, C., Jones, R., Talbot, H., Wu, X., Cheong, K., Beare, R., Buckley, M., and Berman, M. (2006). Measuring the distance of vegetation from powerlines using stereo vision. *ISPRS Journal of Photogrammetry & Remote Sensing*, 60(4):269–283.
- Viswanath, P., Kakadiaris, I. A., and Shah, S. K. (2009). A simplified error model for height estimation using a single camera. In *Proceedings of the 9th IEEE International Workshop on Visual Surveillance*, pages 1259–1266.
- Zhou, J. and Li, B. (2007). Exploiting vertical lines in vision-based navigation for mobile robot platforms. In *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing*, pages 465–468.