

OBJECTS POSITIONING IN WATER SURFACE FROM A SINGLE IMAGE

J. Romeo, J. Aranda

Dpt. Informática y Automática. E.T.S. Informática, UNED
jromeo99@hotmail.com; jaranda@dia.uned.es

G. Pajares

Dpt. Ingeniería del Software e Inteligencia Artificial. Facultad Informática, Universidad Complutense de Madrid
pajares@ucm.es

J.M. de la Cruz

Dpt. Arquitectura Computadores y Automática. Facultad de Ciencias Físicas, Universidad Complutense de Madrid
jmcruz@ucm.es

Abstract

Boats positioning is an important task required for autonomous navigation and control. This paper proposes a new method for estimating real positions (x - y coordinates) of small boats in the 3D scene from a single image. These positions are obtained with respect to a world reference system by the only means of using rows and columns of pixels where the boat is located in the image. A single camera is needed without requiring knowledge of its intrinsic and extrinsic parameters. The proposed approach avoids the use of more complex and sophisticated systems such as GPS. The method is valid for any object on a planar surface with the only constraint that the camera must be on a fixed position. It has been proven with a differential GPS of high precision. The main contribution is made on the computation of both x and y coordinates from a reference system. This method is also valid for objects in other planar surfaces, such as flat fields, crops, indoor floors or horizontal roads.

Keywords: 3D boat positioning, x - y coordinates estimation, depth determination, single camera.

1 INTRODUCTION

There are many depth computation techniques, based on computer vision, to determine the distance of an object in the scene 3D to a reference system. Most of them are focused on the computation of distances without considering the object location. This means that different locations provide the same distance. Figure 1, displays a pedagogical example where the object located at position (x, y) is also placed at the distance d , but this distance is the same for all points in the circumference with radius d and center in the camera.

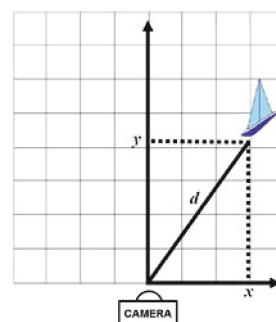


Figure 1: x - y Coordinates and distance d of an object from a reference system whose origin is set on the camera.

Autonomous navigation of unmanned boats requires the knowledge of their locations with respect a reference system for control purposes. The proposed method is specifically designed to provide the required boat 3D locations based only on the computation of their corresponding row and column pixels coordinates on a single image. Motivational research is supported by the above advantages, where the main contribution of this method is the computation of both x and y 3D coordinates once rows and columns of pixels belonging to objects in the image are obtained.

Methods for depth estimation are classified in a field called: 3D Computer Vision. The main goal of this area is the design of strategies to determine the position of objects in the 3D scene. These techniques may be grouped into two categories: active and passive methods [1]. Active techniques apply energy through a structural source (laser, radar, patterns of light) and receive responses through a detector. Based on the nature of the interaction with the surfaces of the objects, the reflected energy is used for recovering the 3D structure in the scene. However, in outdoor environments there are serious limitations and conflicts with variations in object reflectance, color and range. Alternatively, passive

vision techniques do not apply specific structured source of energy to build images avoiding interferences with the environment. This paper is concerned with passive methods. Most works have addressed the topic of 3D determination with passive methods; they are grouped under the generic expression of shape from X, where X can be [2]: *a) stereopsis* [3]; *b) motion*, where the object is moving and the camera is static [4], [5] or the camera moving into the 3D space [6] *c) shading* [7], [8], [9] *d) texture* based on repetition of objects or patterns in the scene [10], [11] *e) focus/defocus*: based on a well-focused object which is mapped exactly on the image plane [12], [13], [14] *f) depth estimation with a single camera*, [15] apply supervised learning; [16] determines relationships between distances of objects in the scene and their corresponding pixels in the image.

a) Stereopsis: based on two cameras separated a base-line, a 3D point in the scene is mapped into different row (r) and column (c) pixel coordinates on each image. A matching procedure is applied to identify the corresponding point in both images. Once the correspondence is established, the 3D pose is obtained by triangulation. Matching is a critical issue in stereovision, which becomes ineffective when large homogeneous areas appear in the scene. This is the case in the proposed application where the water surface displays high texture homogeneity. Depth estimation is commonly performed by stereopsis based techniques [3]. In recent years several modifications of the classical methods have been developed and applied in metrology to determine features of the objects in images, e.g. height, volume or distance. In [17] et al., is proposed a robust method to measure object height based on two cameras, it computes the height using different planes of the image and geometric transformations. On the other hand, there are algorithms that only obtain relative depths and disparity maps without object localization in the scene [18], [19], [20]. Other implementations combine algorithms in order to reduce the computational time required to obtain depth maps [21], [22]. Most of such approaches perform depth and object features estimation under acceptable computation conditions. Nevertheless, they still need intrinsic and extrinsic parameters for both cameras, they don't provide numeric values of the distance between cameras and objects and only have been tested in indoor environments with controlled illumination images.

b) Motion: moving objects produce changes in subsequent images displaying areas which change with respect the background. These methods compute relative motion between camera and 3D scene with small changes. These changes are detected by differential analysis in both spatial and temporal dimensions (optical flow). 3D reconstruction becomes a complex task because the differentiation

requires small movements with few and short features in objects [23]. In our scenario, boats movements are not necessarily small. One way to acquire the 3D data is to determine the direction of translation through approximate motion parallax. Depths can be obtained by least-squares estimations from the optical flow [24]. Because of the above mentioned continuous displacements, recovery of motions and 3D structure is highly affected by this effect, becoming ineffective in our outdoor environments for boats location.

c) Shading: applies patterns of lights and the corresponding shades to infer the shape. It is required the previous knowledge of angles of incidence of the light and also a surface shape estimation. Shading analysis fails when changes in the reflectance occur; this is quite common in outdoor environments. Objects on the water surface, at certain distances, do not display sufficient shades for 3D recovering and do not either show known surface shapes. Moreover, only relative depths are obtained without relation to the world reference system.

d) Texture: methods based on texture analysis exploit the existence of regular repetitions of an element or pattern, called surface texel, in the surface of the structures in the image. Thus, based on the distortion of these texels, once they are mapped over the image, the 3D structure can be inferred by analyzing the slant of such texels when they are mutually compared or by analyzing the distortion of patterns due to perspective projection, e.g.: circumferences are mapped as ellipses. This last assumption requires that texture features are constant within regions of uniform texture. In our water scenario, repetitions of objects or constant features do not appear and the water surface is large with only small boats present displaying several faces, thus texture based-approaches become useless. Additionally, it is required knowledge of the camera's intrinsic parameters.

e) Focus/defocus: it is based on the assumption that well-focused objects in the image are mapped on the image plane, thus if we know the focal length, a simple triangulation allows determining the 3D position. This approach requires adjusting the focal length to determine well focused areas, but in homogeneous surfaces this becomes impractical. Also this requires a motorized system for lens adjusting requiring special calibration to achieve sufficient accuracies for recovering the 3D structure [25]. Different techniques are available for computing measures of focusing, most of them based on edge sharpness. In our scenario the large water surfaces do not display edges and there are not parts to be checked for focusing. Moreover, boats objects, placed at relatively high distances, do not display edges which can be detected with the required precision for focusing purposes.

f) Depth estimation with a single camera: several approaches under different point of view can be considered here. In [15] a supervised learning-based method is applied, where relationships are established between structures in the image and depths in the scene based on the availability of ground-truths. These relations are later applied for new incoming images from similar scenarios. In our scenario, water planar surface appears as a large structure at very different levels of depth, this makes unfeasible the application of this approach. [26] and [27] consider vanishing points in the image and a reference plane in the scene to compute the projection matrix, from which object's heights are derived. In the same way [28] proposed to measure distances based on a single camera with a specific orientation that is capable to estimate the distance between the object and the camera using specific calibration and trigonometric identities. These approaches are based on the previous knowledge of heights in real objects. In [29] depths are computed based on camera displacements, where a relation is established between camera displacement and the variation of pixels in the image for the object under consideration. In the above approaches objects remain still during depth computation and located to relatively small distances. This differs from our application where boats are moving and also are located at distances relatively high. [16] obtains samples of reference objects from different images. Positions of the reference objects are known in the 3D scene and also pixel y -coordinate in the captured images. A n -degree polynomial is estimated, which describes the y -coordinate in the 3D scene, as a function of pixels where objects are located in the image.

The aim of this study is to present a general method valid for positioning boats in water planar surfaces at relatively large distances (up to around 150 meters). From the point of view of texture image analysis, water surfaces display a high degree of homogeneity. Because of the nature of our problem and based on limitations of above strategies, we propose a method based on a single image which is independent of extrinsic and intrinsic camera parameters, so that calibration problems in this complex scenario can be avoided. Some of the strategies described above provide only relative depths of objects in the scene in relation to other objects, but not x - y 3D locations required for boat positioning. The closest approach to the one proposed in this paper is the method proposed on [16], but insufficient for positioning as it only provides the y -coordinate. We also estimate mathematical functions to establish relations between real coordinates and pixels of objects in the image. These functions, not polynomials, allow estimating, with sufficient accuracy for control purposes, both x and y coordinates involving pixel rows and columns

of objects in the image. This passive method achieves equivalent positioning results to the ones provided by GPS active system, which is used for comparative purposes.

A crucial issue required for object location relies on its identification in the image, out of the scope in this work. Important efforts on this regard have been carried out in similar environments, as the ones reported in [31] based on semantic image segmentation and applying high-level knowledge such as sky is above the water and other considerations.

This paper is organized as follows: section 2 describes the proposed method and details the procedure followed to derive relations between spatial coordinates and pixel rows and columns. Section 3 provides functions finally obtained and errors with respect differential GPS. Finally, section 4 contains conclusions and future trends.

2 METHODS DESCRIPTION

We propose a method to estimate the real 3D x - y coordinates in a scene (z is constant in water surface) corresponding to a pixel in the image by knowing only the row and the column where the pixel is placed in the image. The coordinates are estimated with respect to a world reference system OXYZ, that in our case, we set on the camera with the origin O coincident with the center of the image plane, the Y axis coincident with the optical axis in the camera and perpendicular to the image plane, the Z axis is perpendicular to the ground and it is not needed for this method, Figure 2. We have tested this method in a reservoir with images of a boat acquired from a camera placed on the shore at a certain height from water surface. Initially, to obtain the parameters involved in the mathematical functions to be estimated, a manned boat was equipped with a differential GPS so that its real position (GPS coordinates) was known at every time. The images captured for this estimation were taken while the boat was moving on different positions inside the field of view of the camera and its GPS coordinates were logged.

Since GPS coordinates are not the same as coordinates from our system of reference, we have to make a change of coordinates from GPS axes to ours, which is placed on the camera. From now on when we refer and show GPS coordinates we assume this change of coordinates done.

The proposed approach consists of two stages:

- a) Estimation of two mathematical functions linking pixels rows (r) and columns (c) in the image and real 3D x - y coordinates in the scene.
- b) Once the above two functions are estimated its positioning on every new image taken is provided by

the two estimated functions and the results are compared with the GPS coordinates also logged.



Figure 2 (a) Camera placed on the shore of the reservoir; (b) An image of the 3D scene as it is seen by the camera.

The following two requirements are needed to apply the proposed method:

- a) The camera must be fixed permanently, during both stages.
- b) Objects must be placed on planar surfaces. In our approach this does not represent any inconvenience since the objects are boats placed on the water which is always a planar surface.

This method is also valid for positioning different objects in planar surfaces. Based on this fact, we exploit the following two properties derived from the perspective projection, which is the one provided by the optical system attached to the camera:

a) *The lowest the object is located in an image the closer it is to the observer.* In other words, there is a relation between the row of the image where the object is located and the distance to the observer. This relation is modeled as a mathematical function which links the row of the boat in the image with the real 3D y -coordinate. It is to be estimated as explained below.

b) *The closer an object is to the observer the wider it seems in the image.* That is, horizontal dimensions of an object in an image depend not only on the real dimensions of the object themselves but also on how far the object is to the observer, i.e. a big object which is far away from the observer may seem smaller in the image than a narrow object closer to the observer. This means that the real x -coordinate of an object depends on the column but also on the row where the object appears in the image. This second relation is the other mathematical function to be estimated, as explained below.

Considering these two facts, all we have to do is to find these two mathematical relations between the position of an object in an image, identified by its row, r and column, c and its real position (x - y coordinates) in the scene. Figure 3a displays a boat on the image identified as the pixel placed in the row and column (r, c) in the image. Figure 3b identifies the real coordinates (x, y) in the scene.

Since y -coordinate depends only on the row (r), but x -coordinate depends on the row r and also on the column c , we have to consider two cases separately to find these dependencies: one for the Y -axis and another for the X -axis

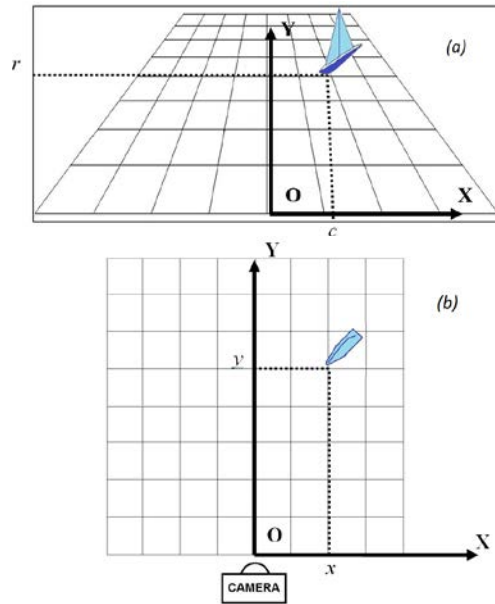
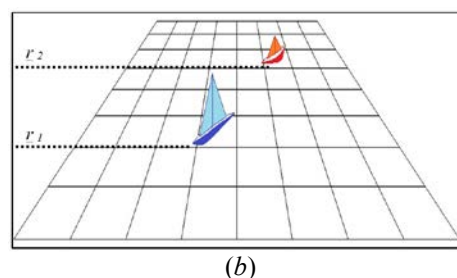


Figure 3 (a): Image acquired by the camera with row, r and column, c ; (b): real coordinates of the boat in the scene with its x - y coordinates with respect the origin O .

Y coordinate, which represents distance to the X -axis, only depends on the row where the pixel is located, that is, the closer the object is to the observer the lower it appears in the image. From figure 4 we can see that for any two coordinates y_1 and y_2 with their respective rows in the image r_1 and r_2 , if $y_1 < y_2$ then necessarily $r_1 < r_2$. Therefore y -coordinates depend only on the row where they are placed in the image. This relation can be established under the assumption that objects are placed on a planar surface on the scene.



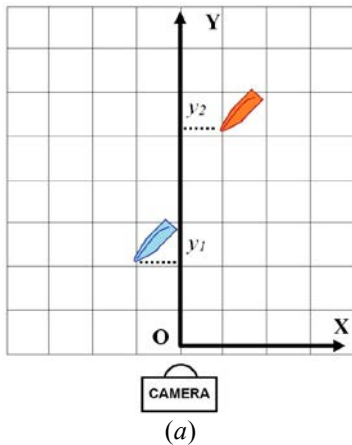


Figure 4 (a): the lowest the object is placed in the image (blue boat, row r1) the closer it is to the observer (blue boat, row r1 (b)).

Now the goal is to find the mathematical function that relates each coordinate y with the corresponding row r . With such a purpose, we take different images of the boat in different positions. These positions are chosen so that they cover the full surface to be analyzed, i.e. with distances from 0 to 150 m from the camera, 150 m is the maximum distance considered in our experiments imposed by the resolution of the camera and the height from the water at which it is placed.

The real 3D position of the boat, y coordinate, is known thanks to the GPS onboard the boat and the row of the boat in the image is obtained directly from the images by locating a marker point in the lowest pixel of the image belonging to the boat. Twelve images were enough for estimating the mathematical functions. Table 1 displays for each image in column 1, positions of the boat at each 3D location, in columns 2 and 4, i.e. y and x coordinates respectively. Columns 3 and 5 displays the corresponding image coordinates in pixels. Column 6 displays what we call ratio (R) concept expressed in pixels/meter, which determines the absolute value of the ratio between each column (c) and its associated x -coordinate.

Relation between y -coordinate and the row r

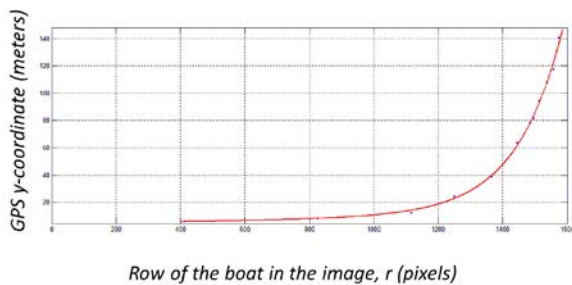


Figure 5 Relation between y -coordinate, distance to the x -axis of the boat, and the row r , where it is located in the image.

Table 1: Twelve images of the boat in different positions, with the y and x coordinates obtained with the GPS, in meters, for each position and their corresponding rows (r) and columns (c) in pixels over the image. The ratio value R is also displayed.

# Image	y -coordinate (m)	row in the image r (pixels)	x -coordinate (m)	column in the image c (pixels)	ratio pixels per meter $R=c/x$ (pixel/m)
1	6,3	413	2,8	742	265,0
2	8,2	825	10	1083	108,3
3	12,3	1.116	-5	-308	61,6
4	24,5	1.249	3,7	146	39,5
5	39,3	1.365	-17,5	-562	32,1
6	63,6	1.445	-9,1	-266	29,2
7	78,3	1.484	39,1	1110	28,4
8	81,8	1.496	49,8	1384	27,8
9	94,5	1.513	47,7	1302	27,3
10	108,0	1.538	-11	-296	26,9
11	117,4	1.558	46,2	1224	26,5
12	140,8	1.574	35,6	935	26,3

Now, the goal is to find the mathematical function describing the relation between the y -coordinates and rows r , expressed as the following model,

$$y = F_y(r) \tag{1}$$

X coordinate of an object, or equivalently distance to the Y -axis, as mentioned before, not only depends on the column where the object is located but also on the row. To understand this fact we take a look at figure 6 that illustrates it. Indeed, in the horizontal axis (X -axis) any given distance, for example 5 meters in row 25 is covered with 673 pixels, however the same distance in a higher row (row 597) is covered with 84 pixels.

Therefore the 3D real x coordinate of the ship is a function of the column and the row where it appears mapped in the image, which can be expressed as follows,

$$x = F_x(c, r) \tag{2}$$

Because of the double dependency of x from c and r , to find how objects width in the X -axis varies with the row of the object in the image we apply a normalisation by computing the ratio R , of pixels per meter in the X -axis. Thus, for every image we calculate the ratio as follows,

$$R = \frac{c}{x} \tag{3}$$

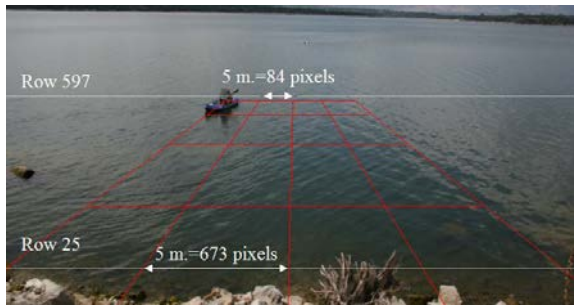


Figure 6 Horizontal distance depends on the column and also on the row of the image.

Figure 7 displays the ratio R against the row r with data in columns 6 and 3 respectively in table 1. Now the goal is to find an analytical function that relates the ratio and the row as follows,

$$R = F_R(r) \tag{4}$$

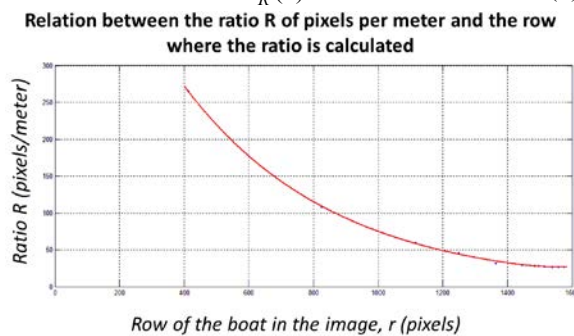


Figure 7 Relation between the ratio R (pixels/meter) and the row of the boat in the image, r .

From equations (3) and (4) we can find the x coordinate of the boat expressed as a function of r and c as follows,

$$x = F_x(c, r) = \frac{c}{F_R(r)} \tag{5}$$

Once functions $F_y(r)$ and $F_x(c, r)$ have been estimated, following the procedure described above, the first stage is finished. Now we can calculate the real x and y coordinates of any pixel in the image by only knowing the row, r and the column, c where it is located in the image. This is the second stage mentioned above.

3 EXPERIMENTAL RESULTS

For estimating the above two functions samples have been taken in Valmayor reservoir, located in Madrid (40° 33' 00.77'' N, 4°03'07.06''O). The experiment was carried out in July 2015 with the following equipment used:

- a) Camera Canon EOS 400.
- b) Differential GPS system, Hemisphere GPS A221, A220.

- c) A 4 meters long rowing boat with a marker point used to identify the boat in the images with a GPS system onboard.
- d) Two radio communicators to synchronize GPS readings with the images taken.

The experiment is summarized in the following steps:

a) A camera is placed on the shore at a certain height from the water surface. The higher it is placed, the more accurate on finding pixels values from the image we will be. This is because if we place it in a high position the water will cover more area in the image than placing it at the water surface level. In our case we placed it on a small hill close to the lake side. Knowing the numeric value of the height is not necessary for calculations.

- b) The boat is moving inside the field of view of the camera.
- c) Different images are taken as the boat moves. For every image acquired, the boat GPS coordinates are also stored. In our case, twelve images were taken and used as samples to fit the functions, table 1.

Applying the method described in section 2 the following analytical functions, with different parameters, were tested for the fit using the procedure cftools [30]: Polynomial, Power, Fourier and Exponential. After different experiments, the best fitting was obtained with the exponential functions given in (6) and (7).

$$y = F_y(r) = 4,6e^{0,000602r} + 0,0023e^{0,006936r} \tag{6}$$

$$x = F_x(c, r) = \frac{c}{650e^{-0,0022r} + 3,0 \times 10^{-8} e^{0,012r}} \tag{7}$$

Functions (6) and (7) were obtained with the following goodness of fit:

SSE: 69.45	SSE: 34.59
R-square: 0.997	R-square: 0.9993
Adjusted R-square: 0.9959	Adjusted R-square: 0.9991
RMSE: 2.946	RMSE: 2.079

After this estimation, we have tested our approach with several images captured from the same scene, which are different from the ones used for the fitting above. For testing purposes we compare y coordinates provided by the GPS and the same coordinates calculated with $F_y(r)$. Table 2 displays the accuracy of our method in 10 images with the absolute and relative errors shown in Table 2.

On the same way, Table 3 shows the values of the x coordinates provided by the GPS, the estimated coordinates using $F_x(c, r)$ and the absolute and relative errors.

Table 2: GPS (real) y -coordinates, our estimated y -coordinates and their absolute and relative error.

Object in the image number	y -coordinate given by the GPS (m)	y -coordinate estimated with the method (m)	Absolute error	Relative error (%)
13	6,08	6,16	0,08	1,32
14	7,12	7,06	0,06	0,84
15	7,58	7,44	0,14	1,85
16	9,50	9,70	0,20	2,11
17	17,41	17,83	0,42	2,41
18	22,18	21,61	0,57	2,57
19	67,13	69,48	2,35	3,50
20	82,88	86,28	3,40	4,10
21	105,16	100,01	5,15	4,90
22	133,92	126,96	6,96	5,20
Mean			1,93	2,88

Table 3: GPS (real) x -coordinates, estimated x -coordinates and their absolute and relative error.

Object in the image number	x -coordinate given by the GPS (m)	x -coordinate estimated with the method (m)	Absolute error (m)	Relative error (%)
13	-1,34	-1,32	0,01	1,1
14	-1,69	-1,72	0,02	1,4
15	-2,05	-2,09	0,04	1,9
16	-1,55	-1,58	0,03	2,1
17	0,52	0,51	0,01	2,4
18	1,14	1,10	0,03	2,9
19	12,22	12,55	0,33	2,7
20	24,56	25,32	0,76	3,1
21	33,75	32,64	1,11	3,3
22	44,71	43,41	1,30	2,9
Mean			0,37	2,38

From table 2 and table 3 we can see how errors increase as y coordinate increase. This is because as we increase distance of the object in the image, ratio R decreases which means that we have less pixels resolution for the same distance, as explained in Figure 6.

Error may be reduced by increasing the height of the camera and modifying the pitch angle so that our scenario covers as much area in the image as possible (Figure 8). This is because the more the rows of pixels the scenario covers in the image the more the number of rows to assign to any distance.

Figure 8 is a computer simulation of a four by four meters grid. Camera has been placed at 4 meters away from the grid. Focal length is 10 mm and camera height and pitch angle on every image, from left to right, are: (a) 0,5 m 0 degrees and (b) 2 m 23

degrees respectively. Notice that as height increases we have more pixels per row for the same distance.

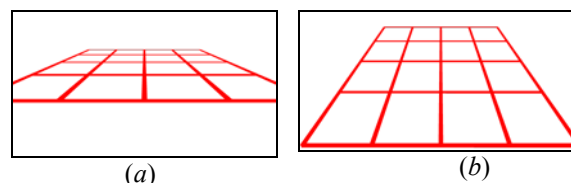


Figure 8 Different images of the same scenario taken from different heights and pitch angles of the camera.

This method has also been tested with images of objects in long indoor corridors (40 meters long) and small rooms (3 meters long) with better results due to the accuracy in the measures of the samples that are needed to fit the analytical functions.

4 CONCLUSIONS

This paper has presented an alternative method not only for depth estimation but also for real x - y coordinates estimation for planar surfaces. It is achieved with a single camera whose intrinsic and extrinsic parameters are not needed. The only requirement is a previous data collection to find the analytical functions. Once functions are found the camera must be fixed. A number of samples is needed to find the functions $F_y(r)$ and $F_x(c,r)$. The number of samples needed will depend on the accuracy of the measures, the resolution of the images and the field of view we want to cover.

This method, based on a single camera, can be easily used in other applications for locating objects placed on planar surfaces i.e.: sea, flat and crop fields, indoor floors, planar roads, etc.

Future works will be comparing these results with those obtained using its intrinsic and extrinsic parameters.

Acknowledgements

The research leading to these results has been funded under by the project DPI2013-46665-C02-01 from the Ministerio de Educación y Ciencia of Spain within the Plan Nacional of I+D+i.

References

- [1] Kumar-Mada, S., Smith, M.L. Smith, L.N. and Midha, S. (2002). An overview of Passive and Active Vision Techniques for Hand-Held 3D data acquisition. Optical Metrology, Imaging and Machine Vision (A. Shearer, F.D. Murtagh, J. Mahon, P. F. Whelan, Eds.). Proc. of SPIE, vol. 4877, pp. 16-27.
- [2] Trucco, E. and Verri, A (1998). Introductory Techniques for 3D Computer Vision. Prentice Hall, Upper Saddle River, NJ.

- [3] Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1), 7–42.
- [4] Buch, N., Orwell, J., Velastin, S. (2009). 3D Extended Histogram of Oriented Gradients (3DHOG) for Classification of Road Users in Urban Scenes. In Proc. of BMVC, September 2009.
- [5] Toshev, A., Makadia, A., Daniilidis, K. (2009). Shape-based object recognition in videos using 3D synthetic object models. In Proc. of CVPR, pp. 288-295, June 2009.
- [6] Leibe, B., Cornelis N., Cornelis, K., Van Gool, L. (2007). Dynamic 3D Scene Analysis from a Moving Vehicle. In Proc. of CVPR, June 2007
- [7] Horn, B. (1986). *Robot Vision*, MIT Press.
- [8] Zhang, R., Tsai, P. S., Cryer, J.E., Shah, M. (1999). Shape from shading: a survey. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(8), 690-706.
- [9] Prados E., Faugeras O. (2006). Shape from shading, handbook of mathematical models in computer vision, Springer pp. 375-388.
- [10] Payet, N. and Todorovic, S. (2011). Scene Shape from Texture of Objects. In Proc. IEEE Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, pp.2017-2024.
- [11] Lobay, A. and Forsyth, D.A. (2006). Shape from texture without boundaries. *Int. J. Computer Vision*, 67(1):71–91, 2006.
- [12] Krotkov, E. (1987). Focusing. *International Journal of Computer Vision*, 1(3): 223-237.
- [13] Subbarao, M., Choi, T. and Nikzad, A. (1993). Focusing techniques. *Optical Engineering*, 32(11): 2824-2836.
- [14] Subbarao, M., Surya, G. (1994). Depth from defocus: A spatial domain approach. *International Journal of Computer Vision*, 13(3), 271-294.
- [15] Saxena, A., Chung, S.H. and Ng, A. Y. (2008). 3-D Depth Reconstruction from a Single Still Image. *Int. J. Computer Vision*, 76, 53-69.
- [16] Kaushik, K T. (2012) Formulation of a n-degree polynomial for depth estimation using a single image. Cornell University Library (available <http://www.oalib.com/paper/3623329#.V2TzFY9OI2w> (accessed June 2016).
- [17] Chen, Z., Pears, N., Liang, B., (2006). A method of visual metrology from uncalibrated images. *Pattern Recognition Letters*, 27, 1447-1456.
- [18] Aziz, M., Mertsching, B. (2010). Fast depth saliency from stereo for region-based artificial visual attention. *Lecture Notes in Computer Science*, 6474, 367-378.
- [19] Daolei, W. and Lim, K. (2011). Obtaining depth map from segmented-based stereo matching using graph cuts. *Journal of Visual Communication and Image Representation*, 22, 325-331.
- [20] Agrawal, S., Halder, A., Biswas, S., Sarkar, D., Sarkar, P.P. (2007). A novel method for estimation of distance of an object from camera. *IET-UK International Conference on Information Communication Technology in Electrical Sciences (ICTES)*, pp. 751-753.
- [21] Shan-Shan, C., Wu-heng, Z., Zhi-lin F. (2011). Depth estimation via stereo vision using Burchfield's algorithm. *Proc. IEEE 3rd International Conference on Communication Software and Networks (ICCSN)*, pp. 403-407.
- [22] Birchfield, S., Tomasi, C. (1998). Depth discontinuities by pixel-to-pixel stereo. *Sixth International Conference on Computer Vision*, pp. 1073-1080.
- [23] Ozden, K.E., Schindler, K. and van Gool, L. (2010). Multibody Structure-from-motion in practice. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 32(6), 1134-1141.
- [24] Cantzler, H. (2016). An overview of range images. In *CVOnline: On-Line Compendium of Computer R. Fisher (ed.)*, Available: <http://homepages.inf.ed.ac.uk/rbf/CVonline/> (accessed June 2016).
- [25] Tao, A., Gang, J., Jian, W. (2010). Auto-focusing Technique for Moving Object Based on Image Definition Criterion. *2010 International Conference on Computer Application and System Modeling (ICCSM 2010)*, vol. 2, pp. 37-42.
- [26] Wang, G., Hu, Z., Wu, F., Tsui, H. (2005). Single view metrology from scene constraints. *Image and Vision Computing*, 23, 831-840.
- [27] Criminisi, A., Reid, I., Zisserman, A. (2000). Single view metrology. *International Journal of Computer Vision*, 40 (2), 123-148.
- [28] Fernandes, J. C., Neves, J. A. B. (2006). Angle invariance for distance measurement using a single camera. *IEEE International Symposium on Industrial Electronics*, pp. 676-680.
- [29] Hsu, C., Lu, M., Wang, W., Lu, Y. (2009). Distance measurement based on pixel variation of CCD images. *ISA Transactions*, 48, 389-395.
- [30] Matlab(2012), The Mathworks. <http://www.mathworks.es/>.
- [31] Kristan, M., Sulic Kenk, V., Kovacic, S., & Pers, J. (2016). Fast image-based obstacle detection from unmanned surface vehicles. *IEEE Trans. on Cybernetics*. 46(3), 641-654.