

Journal of Visual Language and Computing

journal homepage: www.ksiresearch.org/jvlc

Ceiling-Vision-Based Mobile Object Self-Localization: A Composite Framework

Alfredo Cuzzocrea^{a,*}, Luca Camilotti^b and Enzo Mumolo^b

^a*iDEA Lab, University of Calabria, Rende, Italy & LORIA, Nancy, France*

^b*University of Trieste, Trieste, Italy*

ARTICLE INFO

Article History:

Submitted

Revised 6.1.2021

Second Revision 8.1.2021

Accepted 9.7.2021

Keywords:

Localization

Ceiling Landmarks

Mobile Objects

Two-Dimensional Dynamic Programming

ABSTRACT

A composite framework for supporting mobile object self-localization based on ceiling vision is presented in this paper. In particular, we consider the estimation of the position of a mobile object using ceiling landmark images acquired by a low resolution camera placed on a mobile object, being light given by electric lamps with circular holders. The pixels of the images of the light holders on the ceiling are mapped on the image plane of the camera by means of a two dimensional dynamic programming algorithm (2D-DPA). The proposed algorithm estimates the distance from the camera lens to the center of the landmarks using only ceiling vision. Other intelligent tools and solutions further improve the accuracy of our self-localization task. Experiments confirm the benefits of our work.

© 2021 KSI Research

1. Introduction

Self-localization of mobile objects is a fundamental requirement for autonomy. Mobile objects can be for example a mobile service robot, a motorized wheelchair, a mobile cart for transporting tasks or similar. Self-localization represents as well a necessary feature to develop systems able to perform autonomous movements such as navigation tasks. Self-localization is based upon reliable information coming from sensor devices situated on the mobile objects. There are many sensors available for that purpose. The early devices for positioning are rotary encoders. If the encoders are connected to wheels or legs movement actuators, relative movements of the mobile object during its path [3] can be measured. Then, mobile object positioning can be obtained with dead-reckoning approaches. Dead reckoning [3] is still widely used for mobile robot positioning estimation. It is also true that dead-reckoning is quite unreliable for long navigation tasks, because of accumulated error problems.

*This research has been made in the context of the Excellence Chair in Computer Engineering – Big Data Management and Analytics at LORIA, Nancy, France

 alfredo.cuzzocrea@unical.it (Alfredo Cuzzocrea);
camilotti@units.it (L. Camilotti); mumolo@units.it (E. Mumolo)

ORCID(s):

Other popular sensor for self-localization are laser or sonar based range finders and inertial measurement devices. In outside scenarios the most popular approaches are based on Global Positioning System (GPS). Due to the importance of self-localization, many other solutions for indoor environment have been proposed so far with different cost and accuracy characteristics. For example the Ultra Wide Band radio signal indoor localization systems [15], or the Bluetooth-based angle of arrival radio devices [19], or a combination of them. However these systems have serious limitations in cost and reliability, respectively. Another important type of sensors which may be used for cost effective self-localization are the CCD cameras, which require computer vision algorithms such as for example visual odometry, [16]. Mobile objects vision based self-localization is currently an open research field [29] and an increasing number of new methods are continuously proposed. As a matter of fact we have to consider that self-localization of mobile objects requires centimeter-level accuracy and Computer Vision is one of the most cost-effective techniques able to reach that accuracy. Consequently, some surveys of Computer Vision based self-localization techniques appeared recently in the literature, [24].

In this paper we describes a novel Computer Vision

based algorithm for estimating the distance from the camera lens to the center of ceiling landmarks with circular shape using a monocular low cost webcam. From that distance, mobile object localization approaches can be easily developed and a simple example is provided in this paper. The images of the ceiling landmarks are projected on the image plane of the camera. The projection can be analytically described, but the projections distortions, which may arise especially when low cost devices are used, may affect the results. To take into account the projection distortions in order to obtain a better precision of the results, we use an approximation of the two-dimensional dynamic programming (2D-DPA) algorithm [13] which finds a sub-optimal mapping between the image pixels of the ceiling landmarks and the image pixels of the landmarks projected on the camera plane. Since optimum 2D-DPA is NP-complete, however, many approximations have been developed. For example, the 2D-DPA technique described by Levin and Pieraccini in [22] has an exponential complexity in the image size, while Uchida and Sakoe describe in [30] a Dynamic Planar Warping technique with a complexity equal to $O(N^3 9^N)$, N being the image size. Lei and Govindaraju propose in [21] a Dynamic Planar Warping approximation with a complexity of $O(N^6)$. However each approximation has some limitation in terms of continuity of the mapping. In this paper we use an approximation of the optimum 2D-DPA with a complexity of $O(N^4)$ [12] which is implemented on a GPU to obtain real-time performance. When the landmark is far from the camera or if the environments has low lighting, an high quantization noise may arise in acquired images. However the algorithm we describe in this paper is particularly robust against noise due especially to the use of two-dimension DPA.

This paper is organized as follows. Section 2 reports on related work. Section 3 the localization problem is described, and in Section 4 the projection distortion is geometrically described, while in Section 5 the two-dimensional Dynamic Programming approximation is described. In Section 6 the proposed algorithm is sketched and in Section 7 the computer vision algorithms for the detection of landmarks on the image plane are reported. Section 8 sketches a possible global localization approach of the mobile object. In Section 9 we report some experimental comparison of the proposed algorithm with state of the art algorithm. Finally, Section 10 concludes the paper with concluding remarks and a suggestion of future works.

2. Related Work

Many papers on vision-based mobile robot self-localization appeared recently in the literature (e.g., [28, 14, 27, 31]).

Moreover, Avgeris *et al.* describe in [1] a self-localization algorithm for mobile robots that uses cylindrical landmarks resting on the floor and a single pivotal camera with an horizontal angle of view of 30-degree. Each cylindrical landmark has a different color in order to be easily detected by the robot. However, frontal vision could be

occluded by objects or people. Such interference can be avoided by placing the landmarks on the ceiling, so that the camera is tilted toward the ceiling. Ceiling vision has been used by many authors to perform mobile robot localization. One of the early proposals is described in [25] and is based upon a digital mark pattern and a CCD camera. The camera is tilted, so the horizontal distance from the ceiling mark pattern is obtained measuring the ratio between the length and the width of the pattern picture.

Kim and Park [18] acquire ceiling images in a small area with a fish-eye lens camera. Ceiling outlines are detected by means of adaptive binarization and segmentation. Robot pose is obtained after identification of the ceiling region and the determination of the center and the momentum of the region. Lan *et al.* describe in [20] a mobile robot positioning algorithm based on artificial passive landmarks placed on the ceiling and infrared sensors. The landmarks are made of reflective film 2D structures containing dots assigned to unique ID's. The infrared sensors consist of an infrared camera and an infrared LED array. A similar approach is described in [32] where artificial passive reflective landmarks are placed on the ceiling and an infrared camera plus an infrared LED source are used to capture the reflection of the IR light on the landmark for estimating the robot pose.

Wang *et al.* describe in [34] a vision control system which captures ceiling RGB images with a camera placed on the robot, converts the image to HSV color space and uses V channel images to reduce the effect of illumination lamps. The common objects and the straight lines on the ceiling are detected by template matching and used to estimate the robot orientation.

Other Computer Vision based approaches are based on the Free Space Density concept. For example, A. Ribacki *et al.* use an upward facing camera to detect the ceiling boundaries and to estimate the ceiling space density from the current image [26]. Other authors, for example [5, 6] use the ceiling depth images for robot localization. In these approaches self-localization is obtained from Principal Component Analysis of ceiling depth images. Ceiling vision is used by many other authors to perform self-localization of mobile robots. For example Lin *et al.* describe in [23] a visual odometry algorithm based on a monocular camera which points to the ceiling. The algorithm uses several local features detectors for matching the features between two sequential frames of the ceiling.

In addition, it should be considered the emerging integration of these topics with the innovative *big data trend* (e.g., [8, 11, 7, 10, 2]). Here, the main research perspective is to take into account the well-known *3V model* of big data, including *volume*, *velocity* and *variety*.

3. Problem Description

We show in Figure 1 a mobile object in an indoor environment. The movable object is equipped with a camera tilted towards the ceiling at an angle φ . We call h the distance between the camera and the ceiling. Moreover in Fig-

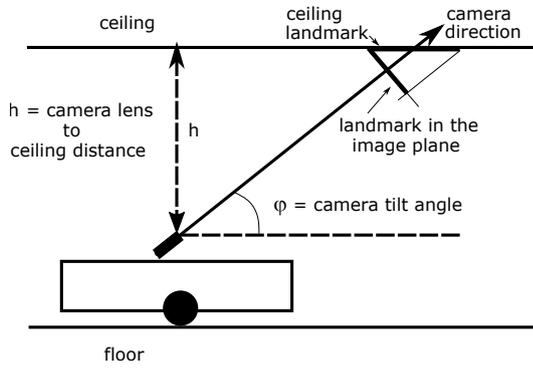


Figure 1: A mobile object with a camera on it, tilted toward the ceiling.

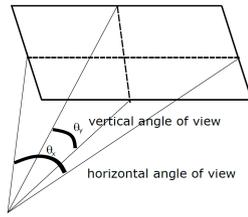


Figure 2: The horizontal and vertical angles of view of the camera.



Figure 3: An example of the circular lamp holder used in this paper.

Figure 2 the horizontal and vertical angles of view of the camera, called θ_x , and θ_y respectively, are highlighted.

The direction towards which the camera is oriented is shown with the 'Camera Direction' arrow. The ceiling landmark is shown in Figure 1 with a segment with a greater thickness and the image plane of the camera is shown with a segment orthogonal to the camera direction. The ceiling landmark is projected to the landmark on the image plane. The visual landmarks positioned on the ceiling used in this approach are the lighting holders as that shown in Figure 3.

We choose landmarks with isotropic shapes because in this way the distortion components due to image rotation can be eliminated. The simplest bidimensional isotropic shape is the circle. As shown in Figure 3, the lines of pixels on the image plane are all parallel to the reference abscissa on the ceiling plane regardless of the angle of the camera with respect to the landmark. It is important to remark that each landmark must be distinguishable from the others and its co-

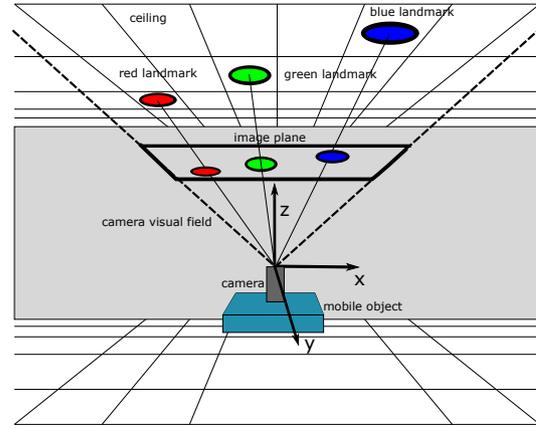


Figure 4: Schematic representation of orthonormal reference system, landmarks and image plane.

ordinate in the global reference system must be known. A schematic representation of a mobile object and some landmarks with the orthonormal reference system centered on the camera lens is shown in Figure 4.

The reference abscissa changes dynamically in relation to the direction of the focal axis. The reference abscissa, in fact, is always normal to the focal axis and at the same time it is parallel to the horizon.

The landmarks must be distinguishable from each other. There are many possible solutions for making the landmarks distinct. A simple possibility is to paint each holder with a different color. More recently, the characteristic frequency of fluorescent lights has been used, for instance in [35]. In this paper we used the simplest solution, namely we painted adjacent lamp holders with different colors.

For this reason the landmarks in Figure 4 are represented with different colors, where for simplicity the three circular landmarks positioned on the ceiling are colored in red, blue and green. Figure 4 shows that the landmarks which fall within the visual field of the camera are projected onto the image plane of the camera. Of course we know in advance the physical position of each landmark in the global reference system. On the other hand the landmark colours can be detected using well known computer vision techniques.

4. Projective Transformations

The projective transformation is the linear transformation of coordinates reported in (1).

$$p' = Tp \quad (1)$$

where p represents a generic point in space expressed in homogeneous coordinates, relative to the orthonormal reference system S described by the quadruple $(O, \hat{i}, \hat{j}, \hat{k})$.

The projected point p' is expressed in coordinates relative to the reference system S' described by the quadruple $(O', \hat{i}', \hat{j}', \hat{k}')$, where $\hat{i}' = \hat{i}$, \hat{j}' has the direction of the segment \overline{MQ} and \hat{k}' has the direction of the normal to the segment \overline{MQ} .

Since p is expressed with the three components (x_p, y_p, z_p) and p' has the three components $(x_{p'}, y_{p'}, z_{p'})$, eq. (1) can be also written as follows:

$$\begin{pmatrix} x_{p'} \\ y_{p'} \\ z_{p'} \end{pmatrix} = T \begin{pmatrix} x_p \\ y_p \\ z_p \end{pmatrix} \quad (2)$$

Such a transformation maintains the properties of collinearity, that is, the points which in S belong to a line, are aligned in a line also in S' . However, projective transformation may not be defined for every point of S , in the sense that some points could be mapped in S' at infinity.

Let us focus on Figure 4 from the left side, that is the $y - z$ plane of the orthonormal reference system which has its origin coincident with the center of the camera lens. This plane is highlighted in Figure 5, where the ceiling is at $z = h$, and the field of view of the camera is shown with points M and E . Let us assume that a landmark falls within the vertical angle of view. Then, the center of the landmark is the point C . On the other hand, if we view Figure 4 from the front side, that is the $x - z$ plane, we obtain the system shown in Figure 6. Of course the camera image plane, which is the plane normal to the focal axis in Figure 4, is shown with the segment $M - Q$ in Figure 5 and segment $G - I$ in Figure 6.

Suppose we fix a point P on the ceiling. If the point falls within the field of view of the camera it is shown as P in Figure 5. Let (p_x, p_y, p_z) , with $p_z = h$, be the coordinates of P . The point P is projected to the image plane of the camera to the point P' , which has coordinates $(x_{p'}, y_{p'}, z_{p'})$. Also the center C of the landmark in Figure 5 is projected to the point C' and the segment $M - E$ is projected to the segment $M - Q$ in the image plane.

In this model, the focal distance of the device or other characteristic parameters are not taken into account. It is in fact a purely ideal model, which has the only purpose of deriving the relations that define the projective transformation from the orthonormal system whose origin coincides with the center of the camera lens to the image plane system. The latter is chosen independently of the characteristics of the camera. With reference to the Figures 5 and 6, we introduce the following geometric variables characteristic of the problem.

- $\Phi = \varphi + \frac{\theta_y}{2} - \frac{\pi}{2}$ (3)

- The distance a from the origin to the barycenter of the landmark projected on the image plane:

$$\begin{aligned} a &= \overline{OC'} = \\ &= \frac{h}{\sin(\varphi)} - h(\tan(\varphi) + \frac{1}{\tan(\varphi)}) \cos(\varphi) \end{aligned} \quad (4)$$

- The abscissa of the point P' on the image plane:

$$\begin{aligned} \frac{b}{2} &= \overline{MC'} = \overline{C'Q} = \\ &= h(\tan(\varphi) + \frac{1}{\tan(\varphi)} \sin(\varphi)) \end{aligned} \quad (5)$$

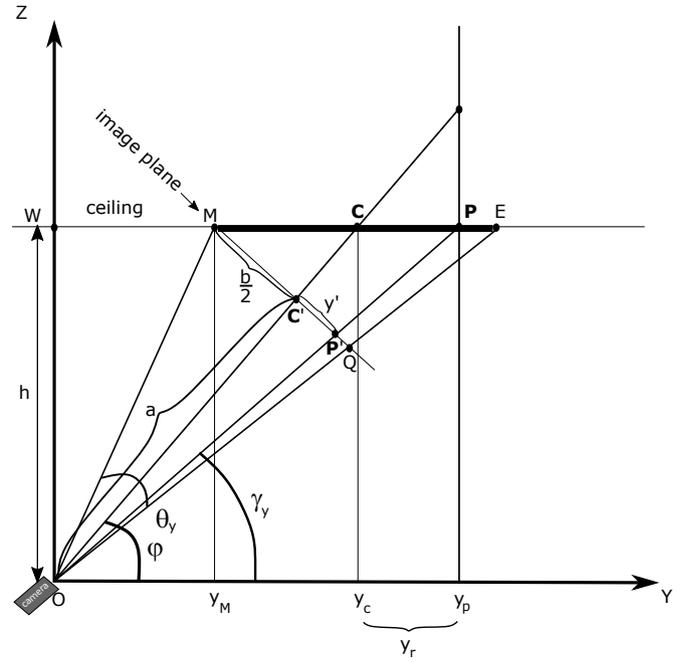


Figure 5: Plane $y - z$ in orthonormal reference system.

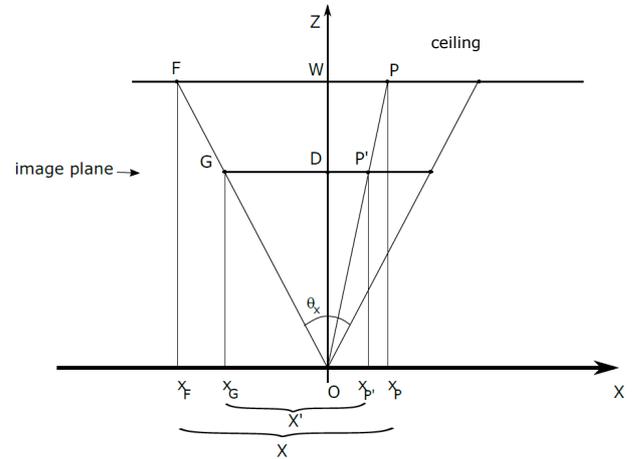


Figure 6: Plane $x - z$ in orthonormal reference system.

Equations (4) and (5) are developed in **Appendix A**. Moreover, we define the following two variables:

$$G = -h(\tan \Phi \tan \varphi + 1) \quad (6)$$

and:

$$F = h(\tan \varphi - h \tan \Phi) \quad (7)$$

We remark that the following considerations are based on three coordinate systems, namely an orthonormal reference system centered on the camera lens, shown in Figure 4, an orthonormal reference system on the image plane and a system on the ceiling plane which is simply translated by h

with respect to that centered on the camera lens. In general, points on the systems centered on the camera lens and on the ceiling are denoted with a capital letter, such as \mathbf{P} , while that on the image plane of the camera are denoted with a capital letter plus an apex such as \mathbf{P}' . In this case, \mathbf{P}' is the \mathbf{P} point projected on the image plane. If we look at the landmark seen from the orthonormal reference system centered on the camera lens, its barycenter is located at (x_c, y_c) . A generic point on the ceiling has coordinate (x, y) and the same point projected on the image plane is (x', y') . The coordinates of a generic point on the landmark is given relative to its barycenter: $(x = x_c + x_r)$ and $(y = y_c + y_r)$. According to Figures 6 and 5 the offsets x_r, y_r are projected to the image plane in x', y' .

Assume now we have an optimum mapping between images. In other words, assume that, having two images A and B , $A = \{a(i, j), i, j = 1, \dots, N\}$ and $B = \{b(u, v), u, v = 1, \dots, M\}$, we can estimate the following mapping function:

$$F(i, j) = \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} x(i, j) \\ y(i, j) \end{bmatrix} \quad (8)$$

which maps each pixel (i, j) of one image to the pixel (u, v) of the other image such that the difference between the two images is minimized, as shown in (9).

$$\min \sum \sum \|a(i, i) - b(u, v)\| \quad (9)$$

where $u = x(i, j)$ and $v = y(i, j)$. Such mapping is performed through a two dimensional Dynamic Programming operation [30]. 2D-DPA is the base of image matching algorithms called Elastic Image Matching. Unfortunately, the Elastic Image Matching operation is NP-complete [17]. For this reason we devise an approximation which reduces the 2D-DPA operation complexity to $O(N^4)$, as described below in Section 5. The barycenter of the landmarks, (x_c, y_c) , are estimated using the following Proposition.

By measuring the abscissa and ordinate (x', y') of a generic point on the landmark projected on the image plane we can estimate the coordinate (x_c, y_c) of the ceiling landmark using the following equations:

$$x_c = \frac{h \cos(\varphi - \gamma_y)(x' - g)}{a \sin(\gamma_y)} + g - x_r \quad (10)$$

$$y_c = \frac{aG + ay_r \tan(\varphi) - (y' - \frac{b}{2})(y_r + F)}{y' - \frac{b}{2} - a \tan(\varphi)} \quad (11)$$

In **Appendix B** we give a sketch of the derivations.

A similar estimation of the coordinates of the landmark barycenter is obtained for other points inside the landmarks. Therefore a sequence of barycenter coordinates x_c, y_c is thus obtained, of which we compute the expected value. The algorithm thus estimates $\mathbf{E}(x_c)$ and $\mathbf{E}(y_c)$ by measuring the values y' and x' of the distorted image on the image plane.

The distance from the camera lens and the landmark in the ceiling reference system is thus the following:

$$d = \sqrt{\mathbf{E}(x_c)^2 + \mathbf{E}(y_c)^2} \quad (12)$$

with reference to Figures 6 and 5, where $C = (x_c, y_c, z_c)$ is the barycenter of the landmark in the reference system (O, i, j, k) . We obtain the sub-optimal correspondence, pixel by pixel, between a reference image and a distorted image by means of approximated two dimensional dynamic programming. Our algorithm therefore uses the deformation of the image to derive the distance of the landmark, i.e. it is intended to determine how the perspective has distorted the image.

The coordinates of the barycenter of the ceiling landmarks are obtained using the coordinate x' measured on the image plane and x_r using the mapping function, and in terms of y' and y_r . Clearly (x_r, y_r) and (x', y') are both known because they are derived from the coordinates of the pixels in the pattern and in the test images respectively. What associates the two pixels is the mapping relationship described in (8) obtained by 2D-DPA.

The characteristic that differentiates the algorithms present in the literature from the one developed in this paper is the statistical character of the obtained estimate. The algorithm based on dynamic programming is able to calculate a position estimate for each single pair of associated pixels from the mapping. The advantage is that a large number of points are used, which contribute to the calculation of the average distance value. This makes the estimate more truthful, especially when the landmark is very distant, which results in a smaller image and a greater quantization error.

5. 2D Dynamic Programming Based Image Mapping Technique (2D-DPA)

For the sake of coherence, we repeat now the mapping considerations summarized above (equations 8 and 9) about images A and B using instead images X and Y . Given two images, $X = \{x(i, j)\}$ and $Y = \{y(u, v)\}$, the mapping of one image to the other is represented by the following operation:

$$D(X, Y) = \min \sum_{i=1}^N \sum_{j=1}^N \|x(i, j) - y(u, v)\| \quad (13)$$

where $u = x(i, j)$, $v = y(i, j)$ is the mapping function between the pixels of X and Y . The quantity $D(X, Y)$ gives a distance between the image X and the optimally deformed Y , the optimal warping function $x(i, j)$, $y(i, j)$ gives an interpretation of the image X according to the generation model Y .

Given the i -th row of the X image and the j -th row of the Y images, namely $Y_j = (y_{j,1}, y_{j,2}, \dots, y_{j,N})$, $X_i = (x_{i,1}, x_{i,2}, \dots, x_{i,N})$ respectively, the distance between the two rows is obtained by applying a 1D-DPA [33] for finding a warping among the two rows as described in (14). Here the map M' is, say, over (n, m) coordinates, so that

$$M'_l = ((i_l, n_l), (j_l, m_l)).$$

$$d(X_i, Y_j) = \frac{\min_{M'} \sum_{l=1}^{M'} d(M'_l)}{M'} = \frac{\min_{M'} \sum_{l=1}^{M'} \|x_{i_l, n_l} - y_{j_l, m_l}\|}{2N} \quad (14)$$

Finally, the distance between the two images is obtained by (15). In this case the map $\overline{M'}$ is between all the rows of X and Y . As before, $|\overline{M'}|$ is the length of the path.

$$D(X, Y) = \frac{\min_{M'} \sum_k d(\overline{M'_k})}{|\overline{M'}|} = \frac{\min_{M'} \sum_k d(X_{i_k}, Y_{j_k})}{|\overline{M'}|} = \frac{\min_{M'} \sum_k \frac{\min_{M'_l} \sum_{l=1}^{M'_l} d(M'_l)}{2N}}{2N} = \frac{\min_{M'} \{ \sum_k \min_{M'_l} \sum_{l=1}^{M'_l} \|x_{i_l, n_l} - y_{j_l, m_l}\| \}}{4N^2} \quad (15)$$

Let us assume that the images are of equal size, that is $N \times N$ pixels. Then the length of the optimum path between the two images is equal to $2N$. The local distances in each point of this path is obtained with other 1D-DPA with paths of length $2N$. The total length is the sum of $2N$ along the $2N$ long path, giving $4N^2$ at the denominator. The complexity of the described operation is $O(N^2 N^2) = O(N^4)$ where N is the image size.

6. Proposed Algorithm

The algorithm described in this paper is summarized as follows. The inputs of the algorithm are the two gray-scale images A and B which are the landmarks on the image plane and on the ceiling respectively. We perform the 2D-DPA algorithm on these two images to obtain the mapping function as result. The mapping function is represented with a linked list where each node is the map related to the two pixels. The function $get()$ gives as result the value of the pixel on the image indicated as input and is used to get the values of the two pixels linked by the map on the two landmark images. To decide if the pixel is a landmark pixel or not, we consider their gray levels. The landmarks have a lower values with respect to the environment and thus if the pixel values is less then a threshold, the pixel is a landmark pixel.

7. Computer Vision Approach for Extracting Landmark Images

We briefly summarize in this Section the computer vision operations we did on the image acquired from the ceiling. The problem is to detect from the image plane the isotropic images which represent the landmark. Another operation, which is not reported here, is the identification of

Input: A, B

Output: *distance*

$img = Detect(A);$ \triangleright get the landmark in the image plane
 $id = identify(img);$ \triangleright identify the landmark

$head = 2D-DPA(A, B);$

$ptr = head;$ \triangleright head is the list of mapping function

repeat

$pixA = get(A, ptr);$ \triangleright pixel of A

$pixB = get(B, ptr);$ \triangleright pixel of B

if $(pixA \leq L) \&\& (pixB \leq L)$ **then** \triangleright if the pixels are in the landmark

 Compute x_c, y_c with (10) and (11)

$sum_y + = y_c;$

$sum_x + = x_c;$

 counter++;

$ptr = ptr \rightarrow next;$

until $ptr == NULL$

$y_c = sum_y / counter;$

$x_c = sum_x / counter;$

$distance = \sqrt{x_c^2 + y_c^2};$

return *distance*

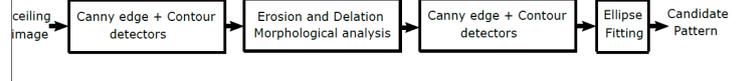


Figure 7: Block diagram of the Computer Vision algorithms.

the landmark. The simplest way is to draw the landmarks with different colors, since the computer vision operations to identify the colors are very simple. There are however many other ways which can be used for the identification, typically based on some type of code drawn inside the landmark. Of course the computer vision operations are slightly more complex than using different colors. More importantly, the computer vision operations to decode drawn codes could need greater camera resolution.

We report in Figure 7 the Computer Vision algorithms we applied on the original image for extraction of isotropic images.

The algorithms are described as follows:

- The acquired image is first transformed in gray-scale, and then its edges are obtained via the Canny's operator, obtaining the Edge image.
- From the Edge image, its contours are extracted, obtaining the Contour1 image.
- The Contour1 image is processed via morphological analysis. More precisely the opening operation with circular structuring element, is applied to Contour1 image in order to eliminate the little Side Dishes. The edges are then extracted again with the Canny opera-

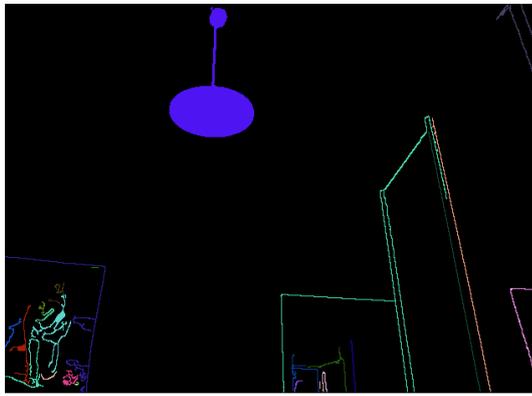


Figure 8: Processed results, with reference to Figure 3.

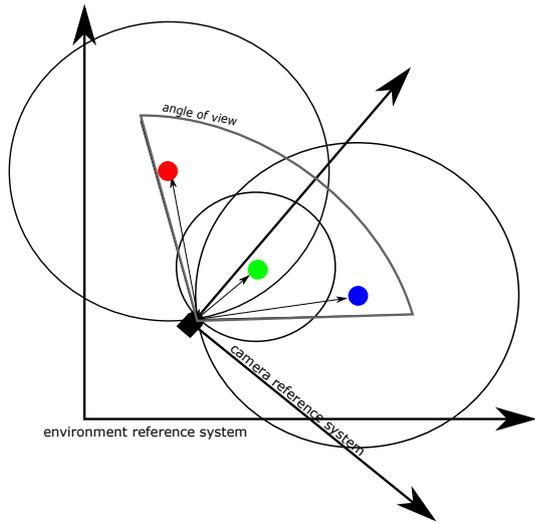


Figure 9: Sketch of a possible localization by trilateration.

tor, and then the contours are extracted again, finally obtaining the Contour2 image.

- Ellipse fitting is applied to Contours2 image. Based on the position and size of the found ellipses, square portions are cut out from original image. Most likely, the landmarks are contained in one of the extracted portions.

The results are shown in Figure 8. These results refer to the input image shown in Figure 3.

8. Localization

The localization of the mobile object is an issue we leave open as starting from distance estimation several possible solution can be developed. However, just to point out a possible simple idea based on trilateration, we report Figure 9.

This figure shows a global reference system which is related to the indoor environment is shown. Another reference system which is rotated and translated with respect to the first one. The origin of second reference system is centered on the camera lens of the mobile object. Note that the $x - y$

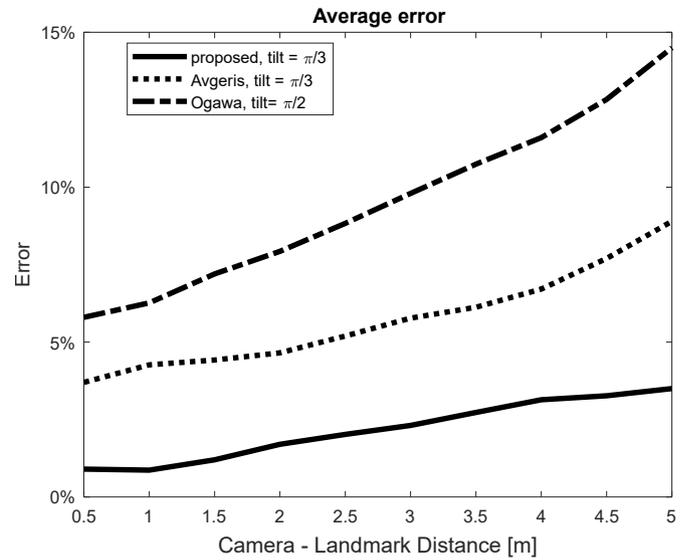


Figure 10: Average errors of the estimated distance.

planes shown in Figure 9 correspond to the ceiling plane. The mobile object identify the landmarks and knows in advance their location coordinate in the global reference system. Our algorithm estimates the distance from the mobile object and the detected landmarks. Therefore, we can think to draw a circle with center on the landmark and radius equal to the estimated distance. If at least three different landmarks are detected, the mobile object can be localized in global reference system.

9. Experimental Results

The experiments has been made using an Intel I7 CPU with 8cores running at 3.07GHz and a memory of 24GB. Then, the two dimensional DPA algorithm has been written in the CUDA framework and executed on a Nvidia Kepler TM GK110 device. A low cost 640×480 webcam is used for image acquisition. In Figure 10 we report the average error of the estimation distance from the camera lens and the barycenter of the landmarks.

As a general consideration regarding these results, if the camera tilt-angle is high (i.e. if the inclination of the optical axis is close to the perpendicular to the ceiling) the error is quite small, but the field of view turns out to be very limited. To take advantage of wider fields of view, higher tilt-angles must be used. In this case, however, the error is higher. Furthermore, if the light in the environment become worse, the average error increases. Our algorithm, however, is quite robust against noise. The curve drawn in Figure 10 with solid line is obtained by the algorithm described in this paper. The curve in the middle is related to the approach developed in 2019 by Avgeris *et al.* and described in [1]. Finally, the higher curve is related to the work proposed by Ogawa *et al.* in [25]. Despite being quite old we include this result because its setting is very similar to this paper (the camera is directed towards the ceiling with a tilt angle equal to 30 de-

grees). The errors are in any cases well above that obtained by all the other algorithms.

10. Final Remarks and Future Work

In this paper we present an algorithm to measure the distance of a mobile object to the lightning lamps used as ceiling landmarks in indoor environment. The algorithm has many attractive features, mainly the accuracy, which is better than many other visual-based algorithms. Also, the distance measurements algorithm is robust against noise. Quantization noise can be high in low lighting condition of the environment and if the distance from landmarks and camera is high. The negative outcome of the algorithm is the high complexity of 2D-DPA which, even if polynomial, can lead to high computational times. In [12], however, we show how the 2D-DPA when implemented on a NVidia Kepler TM GK110 device leads to computation time less than 100 ms, for image size of 100×100 pixels.

This paper naturally opens to the development of localization algorithms based on our distance estimation algorithm. The global localization is in fact under development. Another open important issue is the landmark placement. Finally, the estimation of the orientation of the mobile object is another fundamental problem not addressed in this paper. The use of the characteristic frequencies of fluorescent lamp is an interesting method to identify the landmarks. In this case, adaptive and artificial intelligence metaphors, perhaps inherited by different scientific context (e.g., [4, 9]), may be considered. Future works will be focused on these open points.

Acknowledgements

This research has been partially supported by the French PIA project ‘‘Lorraine Universit  d’Excellence’’, reference ANR-15-IDEX-04-LUE.

Appendix A

Referring to Figure 5, we derive below the geometric variables reported in Section 4.

Consider first Eq. (4),

$$a = \overline{OC'} = \overline{OC} - \overline{C'C} \quad (16)$$

From the right triangle $\triangle OCC_y$, we have:

$$\overline{OC} = \frac{h}{\sin(\varphi)} \quad (17)$$

Moreover, from $\triangle MC'C$, we have:

$$\overline{C'C} = \overline{MC} \cos \varphi = (\overline{WC} - \overline{WM}) \cos \varphi \quad (18)$$

i.e.:

$$\overline{C'C} = \left(\frac{h}{\tan \varphi} + h \tan \Phi \right) \cos \varphi \quad (19)$$

Therefore:

$$a = \frac{h}{\sin(\varphi)} - h \left(\frac{1}{\tan \varphi} + \tan \Phi \right) \cos \varphi \quad (20)$$

Considering Eq. (5), we have:

$$\frac{b}{2} = \overline{MC} \sin \varphi = h \left(\frac{1}{\tan \varphi} + \tan \Phi \right) \sin \varphi \quad (21)$$

Appendix B

We now report a sketch of the derivation of the two propositions reported in Section 4.

Let us start with Eq. (7). Regarding Figure 5, the angle formed by segments \overline{OR} and \overline{OP} is equal to $(\Phi - \gamma_y)$, so:

$$\begin{aligned} \tan(\Phi - \gamma_y) &= \frac{\overline{RP}}{\overline{OR}} = \\ &= \frac{[y \tan \varphi - h(\tan \Phi \tan \varphi + 1)] \cos \varphi}{h(\tan \varphi + y - h \tan \Phi) \cos \varphi} \end{aligned} \quad (22)$$

In addition, to simplifying the $\cos \varphi$, we use the definition of F and G reported above.

$$G = -h(\tan \Phi \tan \varphi + 1) \quad (23)$$

$$F = h(\tan \varphi - h \tan \Phi) \quad (24)$$

Then, we have:

$$\tan(\Phi - \gamma_y) = \frac{y \tan \varphi + G}{y + F} \quad (25)$$

We conclude that:

$$y' = \frac{b}{2} + a \frac{y \tan \varphi + G}{y + F} \quad (26)$$

By setting $y = y_r + y_c$ we obtain the landmark coordinate y_c reported in Eq. (11).

Going now back to Eq. (7), let us consider Figure 6. For lack of space we only state that, according to considerations very similar to that just described, we can conclude that:

$$x' = g + \frac{a \sin \gamma_y (x - g)}{h \cos \varphi - \gamma_y} \quad (27)$$

where $g = \overline{GD} = \overline{DI} = a \tan \frac{\theta_x}{2}$. As we did previously, we substitute $x = x_r + x_c$ in 27 and thus we can obtain x_c , described in Eq. (10).

References

- [1] Avgeris, M., Spatharakis, D., Athanasopoulos, N., Dechouniotis, D., Papavassiliou, S., 2019. Single vision-based self-localization for autonomous robotic agents, in: 2019 7th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), IEEE. pp. 123–129.
- [2] Campan, A., Cuzzocrea, A., Truta, T.M., 2017. Fighting fake news spread in online social networks: Actual trends and future research directions, in: 2017 IEEE International Conference on Big Data (Big Data), IEEE. pp. 4453–4457.

- [3] Campbell, S., O'Mahony, N., Carvalho, A., Krpalkova, L., Riordan, D., Walsh, J., 2020. Where am i? localization techniques for mobile robots a review, in: 2020 6th International Conference on Mechatronics and Robotics Engineering (ICMRE), IEEE. pp. 43–47.
- [4] Cannataro, M., Cuzzocrea, A., Pugliese, A., 2002. Xahm: an adaptive hypermedia model based on xml, in: Proceedings of the 14th international conference on Software engineering and knowledge engineering, pp. 627–634.
- [5] Carreira, F., Calado, J., Cardeira, C., Oliveira, P., 2018. Navigation system for mobile robots using pca-based localization from ceiling depth images: Experimental validation, in: 2018 13th APCA International Conference on Automatic Control and Soft Computing (CONTROLO), IEEE. pp. 159–164.
- [6] Carreira, F., Calado, J.M., Cardeira, C., Oliveira, P., 2015. Enhanced pca-based localization using depth maps with missing data. *Journal of Intelligent & Robotic Systems* 77, 341–360.
- [7] Ceci, M., Cuzzocrea, A., Malerba, D., 2015. Effectively and efficiently supporting roll-up and drill-down olap operations over continuous dimensions via hierarchical clustering. *Journal of Intelligent Information Systems* 44, 309–333.
- [8] Chatzimilioudis, G., Cuzzocrea, A., Gunopulos, D., Mamoulis, N., 2013. A novel distributed framework for optimizing query routing trees in wireless sensor networks via optimal operator placement. *Journal of Computer and System Sciences* 79, 349–368.
- [9] Cuzzocrea, A., 2006. Combining multidimensional user models and knowledge representation and management techniques for making web services knowledge-aware. *Web Intelligence and Agent Systems: An international journal* 4, 289–312.
- [10] Cuzzocrea, A., De Maio, C., Fenza, G., Loia, V., Parente, M., 2016a. Olap analysis of multidimensional tweet streams for supporting advanced analytics, in: Proceedings of the 31st Annual ACM Symposium on Applied Computing, pp. 992–999.
- [11] Cuzzocrea, A., Mansmann, S., 2009. Olap visualization: models, issues, and techniques, in: *Encyclopedia of Data Warehousing and Mining*, Second Edition. IGI Global, pp. 1439–1446.
- [12] Cuzzocrea, A., Mumolo, E., Pirrò, D., Vercelli, G., 2016b. An efficient cuda-based approximate two-dimensional dynamic programming algorithm for advanced computer vision applications, in: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE. pp. 004251–004258.
- [13] Glasbey, C.A., 2009. Two-dimensional generalisations of dynamic programming for image analysis. *Statistics and Computing* 19, 49.
- [14] Gross, H.M., Koenig, A., Boehme, H.J., Schroeter, C., 2002. Vision-based monte carlo self-localization for a mobile service robot acting as shopping assistant in a home store, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE. pp. 256–262.
- [15] He, C., Xia, Y., Yu, C., Jiang, C., 2020a. A multi-hop distributed indoor localization algorithm for ultra-wide-band sensor network, in: 2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV), IEEE. pp. 1335–1340.
- [16] He, M., Zhu, C., Huang, Q., Ren, B., Liu, J., 2020b. A review of monocular visual odometry. *The Visual Computer* 36, 1053–1065.
- [17] Keyser, D., Unger, W., 2003. Elastic image matching is np-complete. *Pattern Recognition Letters* 24, 445–453.
- [18] Kim, Y.G., Park, T.H., 2016. Localization of mobile robots from full detection of ceiling outlines, in: 2016 IEEE International Conference on Information and Automation (ICIA), IEEE. pp. 1515–1520.
- [19] Kumar, G., Gupta, V., Tank, R., 2020. Phase-based angle estimation approach in indoor localization system using bluetooth low energy, in: 2020 International Conference on Smart Electronics and Communication (ICOSEC), IEEE. pp. 904–912.
- [20] Lan, G., Wang, J., Chen, W., 2016. An improved indoor localization system for mobile robots based on landmarks on the ceiling, in: 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO), IEEE. pp. 1395–1400.
- [21] Lei, H., Govindaraju, V., 2004. Direct image matching by dynamic warping, in: 2004 Conference on Computer Vision and Pattern Recognition Workshop, IEEE. pp. 76–76.
- [22] Levin, E., Pieraccini, R., 1992. Dynamic planar warping for optical character recognition, in: *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, IEEE Computer Society. pp. 149–152.
- [23] Lin, Q., Liu, X., Zhang, Z., 2019. Mobile robot self-localization using visual odometry based on ceiling vision, in: 2019 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE. pp. 1435–1439.
- [24] Morar, A., Moldoveanu, A., Mocanu, I., Moldoveanu, F., Radoi, I.E., Asavei, V., Gradinaru, A., Butean, A., 2020. A comprehensive survey of indoor localization methods based on computer vision. *Sensors* 20, 2641.
- [25] Ogawa, Y., Lee, J.H., Mori, S., Takagi, A., Kasuga, C., Hashimoto, H., 1999. The positioning system using the digital mark pattern—the method of measurement of a horizontal distance, in: *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)*, IEEE. pp. 731–741.
- [26] Ribacki, A., Jorge, V.A., Mantelli, M., Maffei, R., Prestes, E., 2018. Vision-based global localization using ceiling space density, in: 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE. pp. 3502–3507.
- [27] Schmitt, M., Rous, M., Matsikis, A., Kraiss, K.F., 1999. Vision-based self-localization of a mobile robot using a virtual environment, in: *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No. 99CH36288C)*, IEEE. pp. 2911–2916.
- [28] Schneegans, S., Vorst, P., Zell, A., 2007. Using rfid snapshots for mobile robot self-localization., in: *EMCR*.
- [29] Shang, W.A., 2019. Survey of mobile robot vision self-localization. *Journal of Automation and Control Engineering* Vol 7.
- [30] Uchida, S., Sakoe, H., 1998. A monotonic and continuous two-dimensional warping based on dynamic programming, in: *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No. 98EX170)*, IEEE. pp. 521–524.
- [31] Utz, H., Neubeck, A., Mayer, G., Kraetzschmar, G., 2002. Improving vision-based self-localization, in: *Robot Soccer World Cup*, Springer. pp. 25–40.
- [32] Vidal, J., Lin, C.Y., 2016. Simple and robust localization system using ceiling landmarks and infrared light, in: 2016 12th IEEE international conference on control and automation (ICCA), IEEE. pp. 583–587.
- [33] Vinotha, K., 2014. Bellman equation in dynamic programming. *Int. J. Comput. Algorithm* 3, 277–279.
- [34] Wang, W., Luo, Z., Song, P., Sheng, S., Rao, Y., Soo, Y.G., Yeong, C.F., Duan, F., 2017. A ceiling feature-based vision control system for a service robot, in: 2017 36th Chinese Control Conference (CCC), IEEE. pp. 6614–6619.
- [35] Zhang, C., Zhang, X., 2018. Visible light localization using conventional light fixtures and smartphones. *IEEE Transactions on Mobile Computing* 18, 2968–2983.