An Innovative Monocular Mobile Objects Self-Localization Approach based on Ceiling Vision

Alfredo Cuzzocrea[1], Luca Camilotti[2], and Enzo Mumolo[2]

[1] Universitá della Calabria, Rende, Calabria, Italy and ICAR-CNR, cuzzocrea@si.dimes.unical.it
[2] Universitá di Trieste, Trieste, Italy, camilotti,mumolo@units.it

Abstract. This study deals with the estimation of the position of a mobile object using ceiling landmarks images acquired by a low resolution camera placed on a mobile object. The mobile object is moving in an indoor environment where light is given by electric lamps with circular holders. The images of the circular holders are projected on the image plane of the camera and are processed by means of computer vision algorithms. The pixels of the images of the light holders on the ceiling are mapped to the pixels of the images of the light holders on the image plane of the camera by means of a two dimensional dynamic programming algorithm (2D-DPA). The projection distortions are thus compensated and this reduces the estimation errors. The algorithm described in this paper estimates the distance from the camera lens to the center of the landmarks using only ceiling vision. Localization can be easily obtain from such distance estimations. The projections are geometrically described and the distance estimation is based on the pixels mapping information obtained by 2D-DPA.

Keywords: localization · ceiling landmarks · mobile object · two dimensional dynamic programming.

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 [1] can be measured. Then, mobile object positioning can be obtained with dead-reckoning approaches. Dead reckoning [1] 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. Other popular sensor devices for self-localization are laser or sonar based range finder devices 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 [2], or the Bluetooth-based angle of arrival radio devices [3], 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 for localization such as for example visual odometry, [20]. Mobile objects vision self-localization is currently an open research field [5] 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 accuracies. Consequently, some surveys of Computer Vision based self-localization techniques appeared recently in the literature, [6].

Many papers on vision-based mobile robot self-localization appeared recently in the literature. For example Avgeris *et al* describe in [15] a self-localization algorithm for mobile robots that uses cylindrical lankmarks resting on the floor and a single pivotal camera with an horizontal angle of view of 30-degree. Each cylindric 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 interferences 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 [17] 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, [7], acquire ceiling images in a small area with a fisheye 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 [8] 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 [9] 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 the IR light on the landmark for estimating the robot pose. Wang *et al.* describe in [18] a vision control system which capture ceiling RGB images with a camera placed on the robot, convert the image to HSV color space and use 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 extract the ceiling boundaries for estimating the ceiling space density from the current image [11]. Other authors, for example [12, 13] 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.

In this paper we describes a novel Computer Vision 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 the 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 is analytical 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 [4] which finds a sub-optimal mapping between the image pixels of the ceiling landmarks and the image plane pixels of the projected landmarks. Since optimum 2D-DPA is NP-complete, in fact, 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 [21] a Dynamic Planar Warping technique with a complexity equal to $O(N^3 9^N)$. Lei and Govindaraju propose in [23] 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 a approximation of the optimum 2D-DPA with a complexity of $O(N^4)$ [19] 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 conditions, an high quantization noise may arise in acquired images. However the algorithm we describe in this paper is particularly robust against noise due expecially to the use of two-dimension DPA.

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

2 Problem Description

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

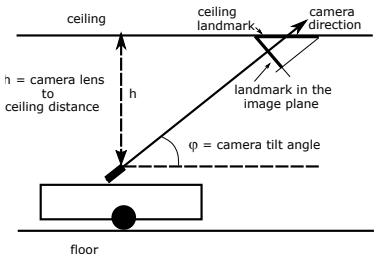

Fig. 1. A mobile object with a camera on it, tilted toward the ceiling.

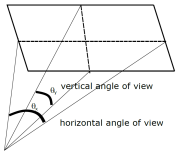

Fig. 2. The horizontal and vertical angles of view of the camera.

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 shown as that shown in Figure 3. We choose landmarks with isotropic shapes on the plane because in this way the distortion components due to image rotation can be eliminated. The simpler 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 coordinate 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

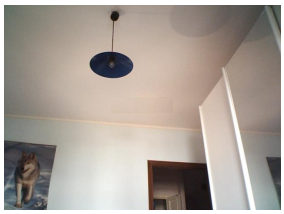

Fig. 3. An example of the circular lamp holder used in this paper

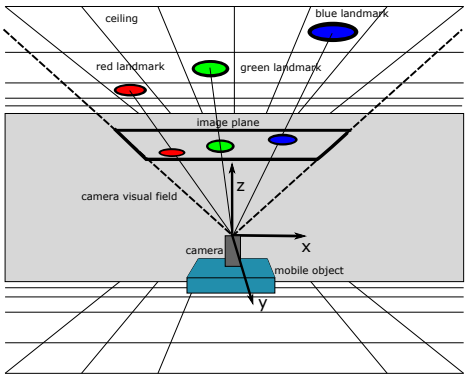

Fig. 4. Schematic representation of orthonormal reference system, landmarks and image plane.

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 [16]. 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.

3 Projective transformations

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

$$p' = Tp \tag{1}$$

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

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

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 view Figure 4 from the left side, that is the $y - z$ plane of the orthonormal reference system which has its origin coinciding 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 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

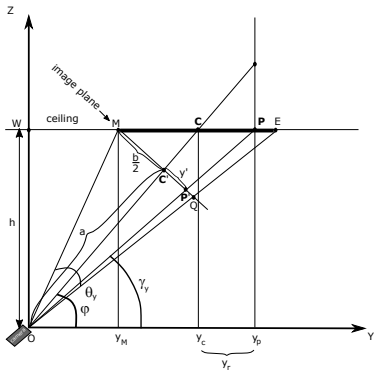

Fig. 5. Plane $y - z$ in orthonormal reference system

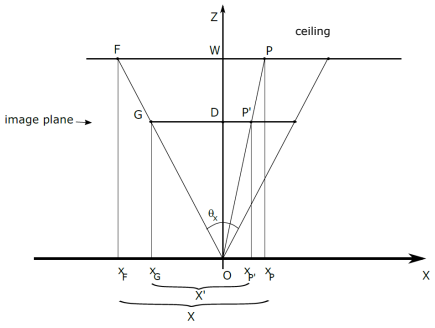

Fig. 6. Plane $x - z$ in orthonormal reference system

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.

–
$$\Phi = \varphi + \frac{\theta_y}{2} - \frac{\pi}{2} \tag{3}$$

– The distance a from the origin ro the barycenter of the landmark projected on the image plane:

$$a = \overline{OC'} = \frac{h}{\sin(\varphi)} - h(\tan(\varphi) + \frac{1}{\tan(\varphi)})\cos(\varphi) \tag{4}$$

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

$$\frac{b}{2} = \overline{MC'} = \overline{C'Q} = h(\tan(\varphi) + \frac{1}{\tan(\varphi)}\sin(\varphi)) \tag{5}$$

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

$$G = -h\left(\tan\Phi\tan\varphi + 1\right) \tag{6}$$

and

$$F = h\left(\tan\varphi - h\tan\Phi\right) \tag{7}$$

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 **P**, while that on the image plane of the camera are denoted with a capital letter plus an apex such as **P'**. In this case, **P'** is the **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, $A = \{a(i,j)|i,j = 1,...,N\}$ and $B = \{b(u,v)|u,v = 1,...,M\}$, we can estimate the mapping function

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

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).

$$min \sum\sum \|a(i,i) - b(u,v)\| \tag{9}$$

where $u = x(i, j)$ and $v = y(i, j)$. Such mapping is performed through a two dimensional Dynamic Programming operation [21]. 2D-DPA is the base of image matching algorithms called Elastic Image Matching. Unfortunately, the Elastic Image Matching operation is NP-complete [26]. For this reason we devise an approximation which reduces the 2D-DPA operation complexity to $O(N^4)$, as described below.

The barycenter of the landmarks, (x_c, y_c), are estimated using the following Proposition.

Proposition 1. *By measuring the abscissa and hordinate (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:*

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

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

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

A different estimation of the coordinates of the landmark barycenter is obtained for all the points $\boldsymbol{P}$ inside the landmarks. A sequence of barycenter coordinates $\boldsymbol{x_c}$, $\boldsymbol{y_c}$ are thus obtained, of which we compute the expected value. The algorithm is thus sequentially divided into two parts: estimation of $\mathbf{E}(\boldsymbol{x_c})$ and $\mathbf{E}(\boldsymbol{y_c})$ by measuring the dimension $\boldsymbol{y'}$ and $\boldsymbol{x'}$ 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:

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

with reference to Figures 6 and 5, where $\boldsymbol{C} = (\boldsymbol{x_c}, \boldsymbol{y_c}, \boldsymbol{z_c})$ is the barycenter of the landmark in the reference system $(\boldsymbol{O}, \boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k})$. We obtain he 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 $\boldsymbol{x'}$ measured on the image plane and $\boldsymbol{x_r}$ using the mapping function, and in terms of $\boldsymbol{y'}$ and $\boldsymbol{y_r}$. Clearly $(\boldsymbol{x_r}, \boldsymbol{y_r})$ and $(\boldsymbol{x'}, \boldsymbol{y'})$ 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.

4 2D Dynamic Programming Based Image Mapping Technique (2D-DPA)

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

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

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

Given the $\boldsymbol{i}-$th row of the $\boldsymbol{X}$ image and the $\boldsymbol{j}-$th row of the $\boldsymbol{Y}$ images, namely $\boldsymbol{Y_j} = (\boldsymbol{y_{j,1}, y_{j,2}, \ldots, y_{j,N}})$, $\boldsymbol{X_i} = (\boldsymbol{x_{i,1}, x_{i,2}, \ldots, x_{i,N}})$ respectively, the distance between the two rows is obtained by applying a 1D-DPA [24] for finding a warping among the two rows as described in (13). Here the map $\boldsymbol{M'}$ is, say, over $\boldsymbol{(n,m)}$ coordinates, so that $\boldsymbol{M'_l} = \boldsymbol{((i_l, n_l), (j_l, m_l))}$.

$$\boldsymbol{d(X_i, Y_j)} = \frac{\min_{\boldsymbol{M'}} \sum_{\boldsymbol{l=1}}^{\boldsymbol{M'}} \boldsymbol{d(M'_l)}}{\boldsymbol{M'}} = \frac{\min_{\boldsymbol{M'}} \sum_{\boldsymbol{l=1}}^{\boldsymbol{M'}} \|\boldsymbol{x_{i_l,n_l}} - \boldsymbol{y_{j_l,m_l}}\|}{\boldsymbol{2N}} \tag{13}$$

Finally, the distance between the two images is obtained by (14). In this case the map $\boldsymbol{\overline{M'}}$ is between all the rows of X and Y. As before, $\boldsymbol{|\overline{M'}|}$ is the length of the path.

$$\boldsymbol{D(X,Y)} = \frac{\min_{\boldsymbol{\overline{M'}}} \sum_{\boldsymbol{k}} \boldsymbol{d(\overline{M'}_k)}}{\boldsymbol{|\overline{M'}|}} =$$

$$= \frac{\min_{\boldsymbol{\overline{M'}}} \sum_{\boldsymbol{k}} \boldsymbol{d(X_{i_k}, Y j_k)}}{\boldsymbol{|\overline{M'}|}} = \frac{\min_{\boldsymbol{\overline{M'}}} \sum_{\boldsymbol{k}} \frac{\min_{\boldsymbol{M'}} \sum_{\boldsymbol{l=1}}^{\boldsymbol{|M'|}} \boldsymbol{d(M'_l)}}{\boldsymbol{2N}}}{\boldsymbol{2N}} =$$

$$= \frac{\min_{\boldsymbol{\overline{M'}}} \{\sum_{\boldsymbol{k}} \min_{\boldsymbol{M'}} \sum_{\boldsymbol{l=1}}^{\boldsymbol{|M'|}} \|\boldsymbol{x_{i,n_l}} - \boldsymbol{y_{j,m_l}}\|\}}{\boldsymbol{4N^2}} \tag{14}$$

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

5 Proposed Algorithm

The algorithm described in this paper is summarized in the following Algorithm. The inputs of the algorithm are the two gray-scale images $\boldsymbol{img_A}$ and $\boldsymbol{img_B}$ which are the landmark 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 $\boldsymbol{get()}$ give 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 thereshold, the pixel is a landmark pixel.

Input:$\boldsymbol{img_A}$, $\boldsymbol{img_B}$
Output: $\boldsymbol{distance}$

img=Detect($\boldsymbol{img_A}$); ▷ get the landmark in the image plane
id=identify(img); ▷ identify the landmark
head=2D-DPA($\boldsymbol{img_A}$, $\boldsymbol{img_B}$);
ptr=head; ▷ head is the list of mapping function
repeat
  $\boldsymbol{pixA = get(img_A, ptr)}$; ▷ pixel of $\boldsymbol{img_A}$
  $\boldsymbol{pixB = get(img_B, ptr)}$; ▷ pixel of $\boldsymbol{img_B}$
  if $\boldsymbol{(pixA \leq L)\&\&(pixB \leq L)}$ **then** ▷ the pixels are in the landmark
    Compute $\boldsymbol{x_c}$, $\boldsymbol{y_c}$ with (10) and (11)
    $\boldsymbol{sum_y += y_c}$;
    $\boldsymbol{sum_x += x_c}$;
    counter++;
  $\boldsymbol{ptr = ptr \rightarrow next}$;
until $\boldsymbol{ptr == NULL}$
$\boldsymbol{y_c = y_sum/counter}$;
$\boldsymbol{x_c = x_sum/counter}$;
$\boldsymbol{distance = \sqrt{x_c^2 + y_c^2}}$;
return $\boldsymbol{distance}$

---

6 Computer vision approaches for extraction of 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 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, tipically 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

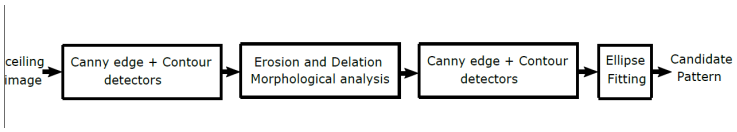

Fig. 7. Block diagram of the Computer Vision algorithms.

as follows:

- The acquired image is first transformed in grayscale, 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 operator, 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.

7 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

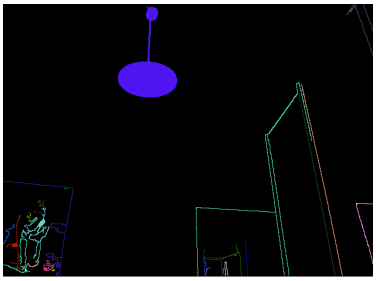

Fig. 8. Processed results, with reference to Figure 3

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

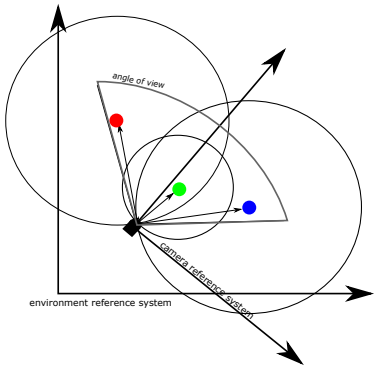

Fig. 9. Sketch of a possible localization by trilateration

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 $\boldsymbol{x-y}$ 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.

8 Experimental Results

The experiments has been made using an Intel I7 CPU with **8**cores running at **3.07**GHz and a memory of **24**GB. 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 $\mathbf{640 \times 480}$ 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

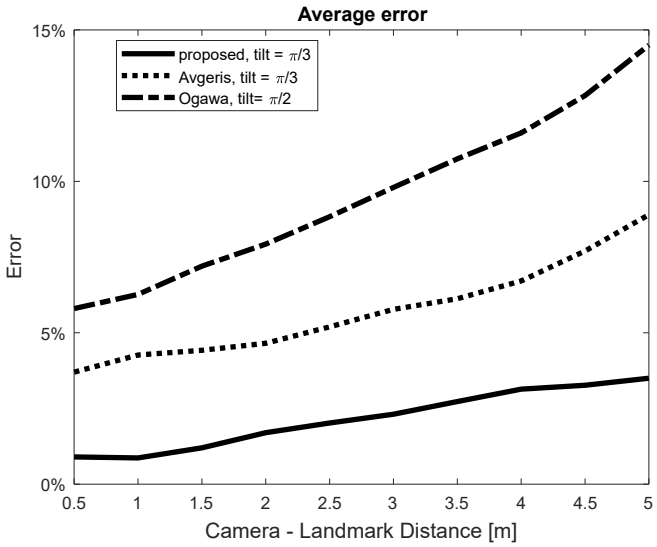

Fig. 10. Average errors of the estimated distance

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 [15]. Finally, the higher curve is related to the work proposed by Ogawa *et al.* in [17]. 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** degrees). The errors are in any cases well above that obtained by all the other algorithms.

9 Final Remarks and future work

In this paper we present an algorithm to measure the distance of a mobile object to the lightings 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 [19], however, we show how the 2D-DPA when implemented on a NVidia Kepler TM GK110 device leads to computation time less then 100 ms, for image size of $\mathbf{100 \times 100}$ 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. Future works will be focused on these open points.

Appendix A

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

Consider first eq. (4). $\boldsymbol{a = \overline{OC'} = \overline{OC} - \overline{C'C}}$. From the right triangle $\boldsymbol{\triangle OCC_y}$ we have $\boldsymbol{\overline{OC} = \frac{h}{\sin(\varphi)}}$. Moreover, from $\boldsymbol{\triangle MC'C}$ we have $\boldsymbol{\overline{C'C} = \overline{MC}\cos\varphi = (\overline{WC} - \overline{WM})\cos\varphi = (\frac{h}{\tan\varphi} + h\tan\Phi)\cos\varphi}$. Therefore

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

.

Considering eq.(5), we have $\boldsymbol{\frac{b}{2} = \overline{MC}\sin\varphi = h\left(\frac{1}{\tan\varphi} + \tan\Phi\right)\sin\varphi}$.

Appendix B

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

Let us start with (7). Regarding Figure **??**, the angle formed by segments $\boldsymbol{\overline{OR}}$ and $\boldsymbol{\overline{OP}}$ is equal to $\boldsymbol{(\Phi - \gamma_y)}$, so $\boldsymbol{\tan(\Phi - \gamma_y) = \frac{\overline{RP}}{\overline{OR}} = \frac{[y\tan\varphi - h(\tan\Phi\tan\varphi + 1)]\cos\varphi}{h(\tan\varphi + y - h\tan\Phi)\cos\varphi}}$. In addition to simplifying the $\boldsymbol{\cos\varphi}$, we use the definition of $\boldsymbol{F}$ and $\boldsymbol{G}$ reported above.

$$\boldsymbol{G = -h(\tan\Phi\tan\varphi + 1)}$$

and

$$\boldsymbol{F = h(\tan\varphi - h\tan\Phi)}$$

. Then we have: $\boldsymbol{\tan(\Phi - \gamma_y) = \frac{y\tan\varphi + G}{y + F}}$. We conclude that $\boldsymbol{y' = \frac{b}{2} + a\frac{y\tan\varphi + G}{y + F}}$ By setting $\boldsymbol{y = y_r + y_c}$ we obtain the landmark coordinate $\boldsymbol{y_c}$ reported in (11).

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

$$x' = g + \frac{a \sin \gamma_y (x - g)}{h \cos \varphi - \gamma_y} \tag{15}$$

where $g = \overline{GD} = \overline{DI} = a \tan \frac{\theta_x}{2}$. As we did previously, we substitute $x = x_r + x_c$ in 15 and thus we can obtain x_c, described in (10).

References

1. Sean Campbell, Niall O' Mahony, Anderson Carvalho, Lenka Krpalkova, Daniel Riordan and Joseph Walsh, Where am I? Localization techniques for Mobile Robots A Review, 6th International Conference on Mechatronics and Robotics Engineering, ICMRE, Barcelona, Spain, February 12-15, pp.43–47, 2020,
2. Chengyang He, Yinqiu Xia, Chengpu Yu and Chaoyang Jiang, A multi-hop distributed indoor localization algorithm for ultra-wide-band sensor network, 16th International Conference on Control, Automation, Robotics and Vision (ICARCV), Shenzhen, China, December 13-15, pp.1335–1340, 2020
3. Gaurav Kumar, Vrinda Gupta and Rahul Tank, Phase-based Angle estimation approach in Indoor Localization system using Bluetooth Low Energy, International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, September 10-12, pp.904–912, 2020
4. Glasbey, C.A., Two-dimensional generalisations of dynamic programming for image analysis, Stat Comput 19, 49, pp. 49–56, 2009.
5. Wei A. Shang, Survey of Nobile Robot Vision Self-localization, Journal of Automation and Control Engineering, Vol.7, No.2, pp.98–101, December 2019
6. Anca Morar, Alin Moldoveanu, Irina Mocanu, Florica Moldoveanu, Ion Emilian Radoi, Victor Asavei, Alexandru Gradinaru and Alexandru Butean, A Comprehensive Survey of Indoor Localization Methods Based on Computer Vision, Sensors, Vol.20, No.9, pp.1–36, 2020
7. Young-Gyu Kim and Tae-Hyoung Park, Localization of mobile robots from full detection of ceiling outlines Proceedings of the IEEE International Conference on Information and Automation, pp.1515–1520, Ningbo, China, August 2016
8. Gongwen Lan, Jingchuan Wang and Weidong Chen, An improved indoor localization system for mobile robots based on landmarks on the ceiling, IEEE International Conference on Robotics and Biomimetics (ROBIO), pp.1395–1399, Qingdao, China, 3-7 Dec. 2016
9. Joel Vidal and Chyi-Yeu Lin, Simple and Robust Localization System Using Ceiling Landmarks and Infrared Light, IEEE International Conference on Control and Automation, pp.583–587, Kathmandu, Nepal, June 1-3, 2016
10. Wenjuan Wang, Zhendong Luo, Peipei Song, Shili Sheng, Yimei Rao, Yew Guan Soo, Che Fai Yeong and Feng Duan, A ceiling feature-based vision control system for a service robot, Proceedings of the 36th Chinese Control Conference, pp.6614–6619, July 26-28, Dalian, China, 2017
11. Arthur Ribacki, Vitor A. M. Jorge, Mathias Mantelli, Renan Maffei and Edson Prestes, Vision-based Global Localization using Ceiling Space Density, IEEE International Conference on Robotics and Automation (ICRA), pp.3502–3507, May 21–25, Brisbane, Australia, 2018

12. Fernando Carreira, João Calado, Carlos Cardeira and Paulo Oliveira, Navigation System for Mobile Robots Using PCA-Based Localization from Ceiling Depth Images: Experimental Validation, 13th APCA International Conference on Automatic Control and Soft Computing, pp.159–164, June 4–6, Ponta Delgada, Azores, Portugal, 2018
13. Fernando Carreira, Joao Manuel Ferreira Calado, Carlos Batista Cardeira, Paulo Jorge Coelho Ramalho Oliveira, Enhanced PCA-Based Localization Using Depth Maps with Missing Data, Journal of Intelligent and Robotic SystemsFebruary 2015
14. Qing Lin, Xiaofeng Liu and Zhihao Zhang, Mobile Robot Self-LocalizationUsing Visual Odometry Based on Ceiling Vision, Symposium Series on Computational Intelligence, pp.1435–1439, Dec.6-9, Xiamen, China, 2019
15. Marios Avgeris , Dimitrios Spatharakis , Nikolaos Athanasopoulos , Dimitrios Dechouniotis and Symeon Papavassiliou, Single Vision-Based Self-Localization for Autonomous Robotic Agents, 7th International Conference on Future Internet of Things and Cloud Workshops, pp.123–129, 26-28 Aug., Istanbul, Turkey, 2019
16. Zhang, C.; Zhang, X., Visible Light Localization Using Conventional Light Fixtures and Smartphones, IEEE Trans. Mob. Comput. 2018, 18, 2968–2983.
17. Yasuo Ogawa and Joo-Ho Lee and Syunji Mori and Akira Takagi and Chie Kasuga and Hideki Hashimoto, The positioning system using the digital mark pattern-the method of measurement of a horizontal distance, IEEE SMC'99 Conference Proceedings, Tokyo, Japan, pp.731–741, 1999
18. Wenjuan Wang and Zhendong Luo and Peipei Song and Shili Sheng and Yimei Rao and Yew Guan Soo and Che Fai Yeong and Feng Duan, A ceiling feature-based vision control system for a service robot, 136th Chinese Control Conference (CCC), Dalian, pp.6614–6619, 2017
19. Alfredo Cuzzocrea and Enzo Mumolo and Daniel Pirro and Gianni Vercelli, An efficient CUDA-based approximate two-dimensional dynamic programming algorithm for advanced computer vision applications, IEEE International Conference on Systems, Man, and Cybernetics, Budapest, Hungary, October 9-12, pp.4251–4258,2016
20. Ming He and Chaozheng Zhu and Qian Huang and Baosen Ren and Jintao Liu, A review of monocular visual odometry, Vis. Comput., 36, 5, pp. 1053–1065, 2020
21. Seiichi Uchida and Hiroaki Sakoe, A monotonic and continuous two-dimensional warping based on dynamic programming, Fourteenth International Conference on Pattern Recognition, Australia, 16-20 August, pp.521–524, 1998
22. Esther Levin and Roberto Pieraccini, Dynamic planar warping for optical character recognition, 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing, San Francisco, California, USA, March 23-26, pp.149–152, 1992
23. Hansheng Lei and Venu Govindaraju, Direct Image Matching by Dynamic Warping, IEEE Conference on Computer Vision and Pattern Recognition Workshops, Washington, DC, USA, June 27 - July 2, 2004,
24. K.Vinotha, Bellman Equation in Dynamic Programming, International Journal of Computing Algorithm, 05, 1, pp = 35–37, 2016
25. Craig Becker and Salas Joaquin and Kentaro Tokusei and Jean-Claude Latombe, Reliable Navigation using Landmarks, IEEE Conference on Robotics and Automation, 2, pp=401–406, 1995
26. Daniel Keysers and Walter Unger, Elastic image matching is NP-complete, Pattern Recognit. Lett., 24, 1-3, pp=445–453, 2003