2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
Daejeon Convention Center
October 9-14, 2016, Daejeon, Korea

Towards an Omnidirectional Catadioptric RGB-D Camera*

José Iglesias, Pedro Mirado, and Rodrigo Ventura

Abstract— **In this paper we address the 3D reconstruction of points, on a non-central catadioptric system, composed by a mirror, a projector, and a perspective camera. The goal of the paper is to propose a framework to build an omnidirectional depth camera, towards an omnidirectional RGB-D camera system. The main contributions are: an efficient technique to project 3D points from the world to an image of a general non-central catadioptric camera; the definition of the template pattern (for both the projector and camera's images); and the matching between the projection of these features to the world and its respective images. The 3D depth is directly recovered using the template matching approach. In conclusion, we apply some filtering techniques to improve the results. To evaluate the proposed framework, we test the method using synthetic data, under different levels and types of noises, proving that the framework is robust to noise and, thus, can be put into practice.**

I. Introduction

RGB-D cameras capture both color and depth images of a scene. Depth maps are images where each pixel contains information about the distance from objects (in the scene) to the respective viewpoint. There are some techniques to obtain these images, which can be either active or passive, depending on whether they require special light sources or not. The most common passive sensor is the stereo camera system. Having multiple views of a scene, depth can be estimated from triangulation, [1]. However, it requires a 3D point to be seen by, at least, two pixels acquired from different viewpoints.

As for active sensors, the most common are time-of-flight cameras (ToF), [2], and structured light sensors, [3]. In the former, a single or several light sources illuminate the scene for a short period of time; the light is reflected by the objects; and its captured by a matrix of detectors, which yield the phase shift between the emitted and reflected beams. These cameras are compact, do not require mechanical parts, and have an high frame rate. Nevertheless, they suffer from the interference of background light (as well as other external light sources), the ambiguity generated by multiple reflections and their high price. The latter consists of projecting a known pattern into the scene. Depth is computed from the shift between the observed and a reference pattern. They are cheaper than ToF cameras, but they are limited to indoor use. The first version of the *Microsoft Kinect* works under this principle. Even though the technology is not disclosed (patent protected), the way depth is estimated can be speculated from some patent applications, e.g. [4] and [5]. In this paper we were inspired by this last latter type of sensor.

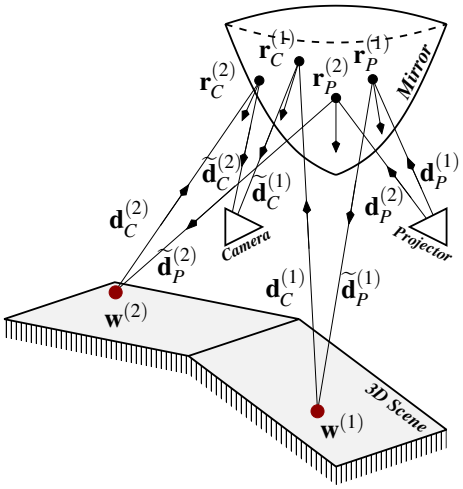


Fig. 1: Representation of the proposed system for the estimation of the depth information: a projector, a RGB camera, and a quadric mirror.

Kinect's high accuracy range data, associated with its robustness to light conditions and speed of measurement, helped increasing its popularity. Between 2010 and 2013, over thousands articles on the use of Kinect were published, with many applications in robotics, such as simultaneous localization and mapping [6], [7], [8] and object perception [9], [10]. The Kinect has become a popular sensor, but it is limited by its field-of-view (43° vertical by 57° horizontal), much smaller than conventional sensors used in robotics, like the laser rangefinder (aprox. 210°). Given this limitation, in this work we propose a theoretical framework to build a structured light omnidirectional RGB-D camera, using a catadioptric systems. The output of the system will be a 3D point cloud, which means that one can use the same algorithms for same robotic's applications that have been used with standard kinect's sensor but, in this case, we would benefit from a wider representation of the environment. The sensor consists of a projector, a RGB cameras, and a quadric mirror (i.e., a mirror described by a quadratic surface).

One of the main challenges in the design of this sensor is related with the reconstruction of the 3D scene (estimation of the D information of the RGB-D image). After this step, if the sensor is calibrated, an additional RGB sensor can be

*This work was supported by the FCT projects with reference [UID/EEA/50009/2013] and [PTDC/EEI-SII/4698/2014]. P. Miraldo was partially funded by the FCT individual grant with reference [SFRH/BPD/111495/2015].

J. Iglesias, P. Miraldo, and R. Ventura are with the "Instituto de Sistemas e Robótica", "Instituto Superior Técnico, Universidade de Lisboa", Portugal. Corresponding Author: pmiraldo@isr.tecnico.ulisboa.pt

used to project the color to the reconstructed scene, setting up all the required data for the RGB-D image. As a result and since this problem was not yet fully addressed in the literature, in this paper we focus our work on the estimation of the D channel, towards an omnidirectional catadioptric RGB-D camera. The proposed system for the 3D scene reconstruction is shown in Fig. 1.

To the best of our knowledge, in the literature there are three different approaches to solve this problem. The first approach consists of having a catadioptric camera and projector in different configurations. Orghidan at [11] proposed such a solution with the camera placed above the projector, which emits a single circle, captured by the camera. However, depth is computed only on that circle. A similar solution was presented by Zhang at [12]. However, the pattern is different consisting of distinctive concentric circles. These allows us to compute depth in more than one plane, but it still does not provide a depth map. On the same line as the previous work, Esperanza at [13] proposed to place the camera and the projector back-to-back with the possibility of projecting different patterns. However, the field-of-view of the sensors do not overlap completely, thus it does not allow for a full depth map. In this paper we aim at proposing a real depth map estimation, given from general 3D points on the image, instead of estimating the reconstruction of circles.

A different way of obtaining an omnidirectional depth information is to use several laser projectors in association with an omnidirectional camera. Zhou at [14] proposed an omnidirectional depth sensor consisting of a perspective camera looking at a square pyramid mirror, which simulates the field-of-view of four virtual cameras. This catadioptric camera is on top of four laser projectors, which combined emit a 360 light stripe. The virtual cameras have a relation of one-to-one with the projectors, giving rise to four structured-light sensors. Each one computes 3D measurements of the laser stripe in its field-of-view. In [15], an omnidirectional 3D camera was proposed. Four diodes emit infrared light, which passes through diffracting optical elements, generating a known pattern of speckles. The pattern is observed by an omnidirectional camera (CCD camera and hyperbolic mirror) and depth is computed by structured light. All of the method mentioned above require more devices that the ones necessary to build an omnidirectional catadioptric camera. In this paper (and contrarily to these method), we aim at proposing a simple omnidirectional camera sensor, more compact and less expensive, because we only use a projector and RGB cameras, looking to the same mirror.

The last type consists on an extension catadioptric for RGB-D cameras, presented in [16]. This consisted of a typical RGB-D sensor placed under two planar mirrors, which allow to use the built-in processing of the camera for the registered RGB-D image. The mirrors split the field-of-view, creating two virtual viewpoints, and hence two virtual RGB-D cameras. Their approach is more similar to ours. The main difference to what we propose is the fact that they are using two planar mirrors (simulating two different RGB-D sensor looking at different place), while we propose the use of a quadric mirror, allowing a wider field of view. Notice that, contrarily to this case, our system will not have the same properties of typical RGB-D sensors, which means that new algorithms and methods for the basic properties of an RGB-D sensor have to be defined.

The main contributions of the paper are:

- The proposal of a novel omnidirectional RGB-D device;
- An efficient technique for the projection 3D points into an image of a general non-central catadioptric camera;
- The definition of a template matching, that will be used for the 3D reconstruction; and
- Matching of the projector image features and image features.

To conclude, we evaluate the proposed system and method for the scene reconstruction using synthetic data, under different levels and types of noises.

II. Proposed Approach

The goal of this paper is to study the effects of having a curved mirror in front of a structured line sensor, useful to have large field of view (Fig. 1 we shows the proposed setup). The reconstruction of the depth information using structure light sensors can be divided into two steps. Firstly, the projector is used to project a known pattern to the world. Then, from the images of the pattern captured by the camera, one can recover the depth and, as a result, reconstruct the world points that were projected.

The projector and the camera are in front of the mirror, which means that one needs to take into account the reflection from both. In Sec. II-A we study the effects of the reflection associated with the projection of the pattern to the 3D scene. In sec. II-B, we address the projection of these patterns into the image of the camera and, in Sec. II-C, we study the matching and reconstruction problem (between the projected patterns and the respective images).

We define the mirror such that any 3D point $\mathbf{r} = [r_1, r_2, r_3]^T$, that is incident with the mirror, satisfies the following relation:

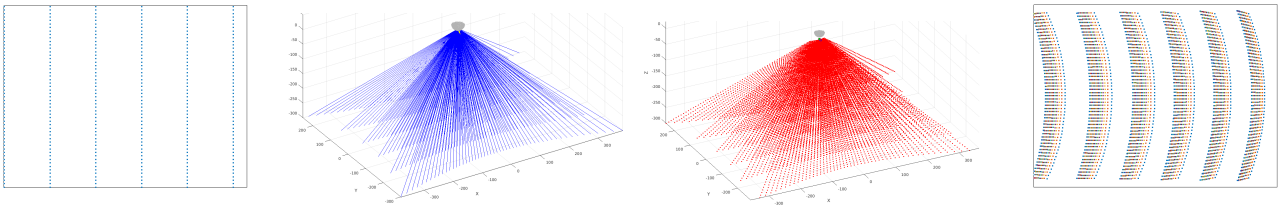
$$Ar_1^2 + Br_2^2 + (r_3 - C)^2 - D = 0,\ r_3 \leq C, \tag{1}$$

where $A, B, C, D \in \mathbb{R}$ are the mirror's parameters. To conclude, we consider the RGB camera, located at $\mathbf{c} = [c_1, c_2, c_3]^T$ and a projector at $\mathbf{p} = [p_1, p_2, p_3]^T$.

A. Pattern projection

The goal of the projector is to project a binary pattern (supose 0's and 1's) to the world. An example of a possible binary pattern is shown in Fig. 2(a), where was used a image size 1348×1048 and the points are separated 500 pixels horizontally and 40 pixels vertically. The spacing between the projected points was designed to improve the matching and reconstruction steps (more detail are given in Sec. II-C).

Let us denote the coordinates $(u_P^{(k)}, v_P^{(k)})$ as the k^{th} non-zero pixel of the projected pattern (see for example Fig.2(a)), $\mathbf{K}_P \in \mathbb{R}^{3\times3}$ be the intrinsic parameter matrix of the projector, and $\mathbf{R}_P$ be its rotation matrix related to the mirror's



(a) Example of an image pattern. (b) Direct projection lines. (c) Discretization of the lines. (d) Projection of the points.

Fig. 2: Example of a possible binary pattern, the template pattern, and its real projection into the image. For this example we used $\mathbf{p} = [0,0,0]^T$, $\mathbf{c} = [0,-7.5,0]^T$, $A = B = 1$, $C = 37.5$, $D = 20^2$ and a specific world. Images with size (image size of 1348×1048).

coordinate system. Then, the direction of the optical ray $\mathbf{d}_P^{(k)} = [d_{P,1}^{(k)}, d_{P,2}^{(k)}, d_{P,3}^{(k)}]^T$ that leaves the projector is given by

$$\mathbf{d}_P^{(k)} = (\mathbf{K}_P \mathbf{R}_P)^{-1} \begin{bmatrix} u_P^{(k)} & v_P^{(k)} & 1 \end{bmatrix}^T. \tag{2}$$

The projection of the pattern onto the mirror's surface $\mathbf{r}_P^{(k)} = [r_{P,1}^{(k)}, r_{P,2}^{(k)}, r_{P,3}^{(k)}]^T$ (for all k), can be written as

$$\mathbf{r}_P^{(k)} = \lambda^{(k)} \mathbf{d}_P^{(k)} + \mathbf{p}, \text{ where } \lambda^{(k)} \in \mathbb{R}^+. \tag{3}$$

To calculate the unknown $\lambda^{(k)}$, we use the condition that any point on the mirror must verify (1). Replacing (3) at (1), one can write

$$a_1 {\lambda^{(k)}}^2 + a_2 \lambda^{(k)} + a_3 = 0, \tag{4}$$

such that

$$a_1 = A {d_{P,1}^{(k)}}^2 + B {d_{P,2}^{(k)}}^2 + {d_{P,3}^{(k)}}^2; \tag{5}$$

$$a_2 = 2(A d_{P,1}^{(k)} p_1 + B d_{P,2}^{(k)} p_2 + d_{P,3}^{(k)}(p_3 - C)); \text{ and} \tag{6}$$

$$a_3 = A p_1^2 + B p_2^2 + (p_1 - C)^2 - D, \tag{7}$$

which give us the two solutions for $\lambda^{(k)}$. Depending on the type of mirror and the projector's position relatively to the mirror, one can easily select the correct $\lambda^{(k)}$[1]. This allow us to compute the intersection point, as written in (3).

The normal to the mirror at $\mathbf{r}_P^{(k)}$ is given by

$$\mathbf{n}_P^{(k)} = [2A r_{P,1}^{(k)}, 2B r_{P,2}^{(k)}, 2(r_{P,3}^{(k)} - C)]^T. \tag{8}$$

Now, from the Snell's law of reflection, we can compute the direction of the reflection ray $\widetilde{\mathbf{d}}_P^{(k)}$ (see Fig. 1), as

$$\widetilde{\mathbf{d}}_P^{(k)} = \mathbf{d}_P^{(k)} - 2\mathbf{n}_P^{(k)} \left(\frac{\mathbf{n}_P^{(k)} \cdot \mathbf{d}_P^{(k)}}{\|\mathbf{n}_P^{(k)}\|^2} \right), \tag{9}$$

which means that the projection line, that projects the projector's pixel k, can be written as

$$\mathbf{l}_p^{(k)} \doteq \alpha^{(k)} \widetilde{\mathbf{d}}_P^{(k)} + \mathbf{r}_P^{(k)}, \text{where } \alpha^{(k)} \in \mathbb{R}^+, \tag{10}$$

see for example Fig. 2(b). For some $\alpha^{(k)}$, this projection line will intersect the world somewhere in a point, that we will denote as $\mathbf{w}^{(k)}$. The aim of this subsection is to compute $\alpha^{(k)}$.

Since the 3D scene is unknown, what we do is to give a set of possible solution for $\alpha^{(k)}$, by discretizing the respective projection lines, such as

$$\{\widehat{\mathbf{w}}^{(k,j)}\} = \{\alpha^{(k,j)} \widetilde{\mathbf{d}}_P^{(k)} + \mathbf{r}_P^{(k)} : j = 1, \ldots, M\} \tag{11}$$

where $\alpha^{(k,j)}$ was previously defined. An example of this discretization is shown in of Fig. 2(c).

Then, the problem can be solved by computing $\mathbf{w}^{(k)}$, from the set of predefined points $\{\widehat{\mathbf{w}}^{(k,j)}\}$. To do this, in this paper, we propose to project this set of points (that belong to the projected line) into the camera's image (see Fig. 1),

$$\{(u_C^{(k,j)}, v_C^{(k,j)})\} = \{\text{Proj}(\widehat{\mathbf{w}}^{(k,j)})\}, \text{ for all } j, \tag{12}$$

where Proj(.) denotes the projection of a 3D point onto the camera's image. We call template matching image (also denoted as $\mathscr{T}(u,v)$) to the image formed with this set of points. An example of the projection of these $\widehat{\mathbf{w}}^{(k,j)}$ into the camera's image is shown in Fig. 2(d). Then, the goal is to use the real image of the unknown $\mathbf{w}^{(k)}$ (acquired by the camera) to do the matching.

In the next subsection, we address the problem of the projection of these set of points onto the camera's image.

B. Images of the patterns

There are some methods to estimate the reflection point in a catadioptric camera system, e.g. [17], [18]. Notice that, so far, there is not a close-form solution for this problem. At [17], Nuno proposed expressing the problem as an intersection of two quadrics, which results in finding the reflection point in quartic curve, using an general iterative technique. On the other hand, Agrawal et al. [18] expressed the problem as a polynomial equation. However, in general (which is the case needed for the problem addressed in this paper), this polynomial equation has degree eight, which means that there are not a close-form solution to this problem. In addition, they consider only mirror axial symmetric mirror.

In this paper, we use known restrictions and propose a novel optimization technique, based on a well known algorithm which we optimize to this problem, improving its computational speed. Using this method, we got an average

[1]For example, considering spheroid or ellipsoid mirrors we would choose the $\lambda^{(k)}$ with the smallest value.

of only 10 iterations for the method to converge to the goal solution. In addition, we also show that, for this specific problem, the proposed optimization technique has a global minimum.

According to the scheme shown at Fig. 1, the points $\widehat{\mathbf{w}}^{(k,j)}$ will appear in the camera's image through a new reflection on the mirror. Let us denote the reflected point on the mirror (that allows the camera to see $\widehat{\mathbf{w}}^{(k,j)}$) as $\widehat{\mathbf{r}}_C^{(k,j)}$. It is known that at the point $\widehat{\mathbf{r}}_C^{(k,j)}$ (which result from the projection of $\widehat{\mathbf{w}}^{(k,j)}$ through the incident direction $\widehat{\mathbf{d}}_C^{(k,j)}$), the projection line reflects on the mirror, according to the same reflection law as defined in (9), and we can define the constraint

$$\left\| \overset{\widehat{=}}{\mathbf{d}}_C^{(k,j)} \times \left(\mathbf{c} - \widehat{\mathbf{r}}_C^{(k,j)} \right) \right\|^2 = 0, \tag{13}$$

where $\overset{\widehat{=}}{\mathbf{d}}_C^{(k,j)}$ is the direction of the reflected ray (see Fig. 1). This happens because the two directions, $\overset{\widehat{=}}{\mathbf{d}}_C^{(k,j)}$ and $(\mathbf{c} - \widehat{\mathbf{r}}_C^{(k,j)})$, must be colinear, which implies that its cross product must be a null vector.

Using the reflection law defined in (9) (but now applied to $\widehat{\mathbf{d}}_C^{(k,j)}$, $\widehat{\mathbf{r}}_C^{(k,j)}$ and $\widehat{\mathbf{n}}_C^{(k,j)}$ – the normal to the mirror at $\widehat{\mathbf{r}}_C^{(k,j)}$) and multiplying both sides of (13) by $\|\widehat{\mathbf{n}}_C^{(k,j)}\|^4$, we get

$$\left\| \|\widehat{\mathbf{n}}_C^{(k,j)}\|^2 \left(\widehat{\mathbf{d}}_C^{(k,j)} - 2\widehat{\mathbf{n}}_C^{(k,j)} \left(\widehat{\mathbf{n}}_C^{(k,j)} \cdot \widehat{\mathbf{d}}_C^{(k,j)} \right) \right) \times \left(\mathbf{c} - \widehat{\mathbf{r}}_C^{(k,j)} \right) \right\|^2 = 0. \tag{14}$$

Since we assume that the camera is calibrated, if we estimate the coordinates of the reflection point on the mirror, one can easily compute the respective image pixel (more details will be given later). Then, the goal is to find $\widehat{\mathbf{r}}_C^{(k,j)}$ such that

$$\widehat{\mathbf{r}}_C^{(k,j)} = \underset{\widehat{\mathbf{x}}^{(k,j)}}{\operatorname{argmin}} f(\widehat{\mathbf{x}}^{(k,j)}), \tag{15}$$

where

$$f(\widehat{\mathbf{x}}^{(k,j)}) = \left\| \|\widehat{\mathbf{n}}_C^{(k,j)}\|^2 \left(\widehat{\mathbf{d}}_C^{(k,j)} - 2\widehat{\mathbf{n}}_C^{(k,j)} \left(\widehat{\mathbf{n}}_C^{(k,j)} \cdot \widehat{\mathbf{d}}_C^{(k,j)} \right) \right) \times \left(\mathbf{c} - \widehat{\mathbf{x}}^{(k,j)} \right) \right\|^2 \tag{16}$$

and the direction $\widehat{\mathbf{d}}_C^{(k,j)}$ and the normal $\widehat{\mathbf{n}}_C^{(k,j)}$ are given by

$$\widehat{\mathbf{d}}_C^{(k,j)} = \frac{\widehat{\mathbf{x}}^{(k,j)} - \widehat{\mathbf{w}}^{(k,j)}}{K_1}, \text{ and} \tag{17}$$

$$\widehat{\mathbf{n}}_C^{(k,j)} = \frac{[2A\widehat{x}_1^{(k,j)}, 2B\widehat{x}_2^{(k,j)}, 2(\widehat{x}_3^{(k,j)} - C)]^T}{K_2}, \tag{18}$$

such that K_1 and K_2 are convenient constants, that smooth the cost function $f(.)$ by making the euclidean norms of $\widehat{\mathbf{d}}_C^{(k,j)}$ and $\widehat{\mathbf{n}}_C^{(k,j)}$ smaller, without turning $f(.)$ into a more complex function (notice that $\widehat{\mathbf{d}}_C^{(k,j)}$ and $\widehat{\mathbf{n}}_C^{(k,j)}$ are vectors that can be defined up to a scale factor).

Knowing that $\widehat{\mathbf{x}}^{(k,j)} = [\widehat{x}_1^{(k,j)}, \widehat{x}_2^{(k,j)}, \widehat{x}_3^{(k,j)}]^T$ has to belong to the mirror, we can write

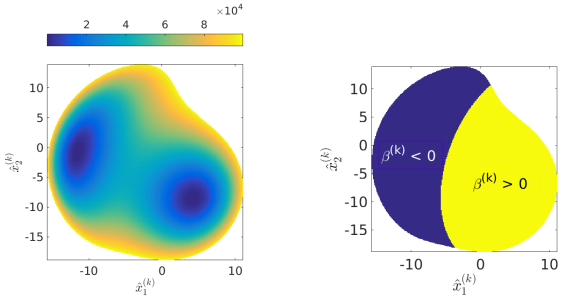
$$\widehat{x}_3^{(k,j)} = C - \sqrt{D - A\widehat{x}_1^{(k,j)\,2} - B\widehat{x}_2^{(k,j)\,2}} \tag{19}$$

and redefine $f(.) \doteq f(\widehat{x}_1^{(k,j)}, \widehat{x}_2^{(k,j)})$.

As said before, (15) hits a minimum if vectors $\overset{\widehat{=}}{\mathbf{d}}_C^{(k,j)}$ and $\mathbf{c} - \widehat{\mathbf{r}}_C^{(k,j)}$ are collinear, which means that they can have the same or opposite directions. In both cases, we can expect to have two solution for: $\widehat{\mathbf{r}}_C^{(k,j)} = \{\widehat{\mathbf{r}}_C^{(k,j,1)}, \widehat{\mathbf{r}}_C^{(k,j,2)}\}$; $\overset{\widehat{=}}{\mathbf{d}}_C^{(k,j)} = \{\overset{\widehat{=}}{\mathbf{d}}_C^{(k,j,1)}, \overset{\widehat{=}}{\mathbf{d}}_C^{(k,j,2)}\}$; and $\widehat{\beta}^{(k,j)} = \{\widehat{\beta}^{(k,j,1)}, \widehat{\beta}^{(k,j,2)}\}$, such that

$$\mathbf{c} = \widehat{\mathbf{r}}_C^{(k,j,i)} + \widehat{\beta}^{(k,j,i)} \overset{\widehat{=}}{\mathbf{d}}_C^{(k,j,i)}, \ \widehat{\beta}^{(k,j,i)}, \in \mathbb{R}, \text{ and } i = 1,2, \tag{20}$$

Since we will have opposite directions for both $\overset{\widehat{=}}{\mathbf{d}}_C^{(k,j,1)}$ and $\overset{\widehat{=}}{\mathbf{d}}_C^{(k,j,2)}$, $\operatorname{sign}(\widehat{\beta}^{(k,j,1)}) = -\operatorname{sign}(\widehat{\beta}^{(k,j,2)})$. This means that, from the two solutions of $f(\widehat{x}_1^{(k,j)}, \widehat{x}_2^{(k,j)}) = 0$, we can only consider the one in which $\widehat{\beta}^{(k,j)} > 0$, which corresponds to the one with the short's ray path (Fermat's principle).



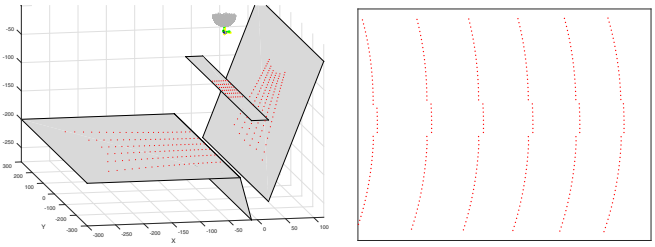
(a) Objective function $f(.)$ colormap.

(b) $\operatorname{sign}(\widehat{\beta}^{(k,j)})$ colormap.

Fig. 3: Example of the variation of the values of the objective function (15), considering the parameters $A = 1.59$, $B = 1$, $C = 35$, $D = 25^2$.

To illustrate this problem, lets consider an example. Fig. 3(a) shows the color map of $f(.)$ for the optimization parameters $\widehat{x}_1^{(k,j)}$ and $\widehat{x}_2^{(k,j)}$. All the possible solutions for the optimization parameters, which means $\widehat{x}_1^{(k,j)}$ and $\widehat{x}_2^{(k,j)}$ that hit the mirror, are shown in this figure. The two minimums mentioned in the previous paragraph are clearly visible. On the other hand, in Fig. 3(b) it is shown the sign of $\widehat{\beta}^{(k,j)}$, of each of the $(\widehat{x}_1^{(k,j)}, \widehat{x}_2^{(k,j)})$ considered. As it can be easily seen from these figures, the desired solution is the one on the right, since it is the one which has $\widehat{\beta}^{(k,j)} > 0$. The separation between the two regions ($\widehat{\beta}^{(k,j)} > 0$ and $\widehat{\beta}^{(k,j)} < 0$) coincides with a local maxima barrier of $f(.)$ and, with that information, one can conclude that any iterative method initialized with a point such that the initial $\widehat{\beta}^{(k,j)} > 0$ will converge to the desired solution.

In order to solve this optimization problem, a gradient descent method with adaptive step size [19], optimized for our problem, was developed. The (iterative) gradient method is commonly used in minimization of unconstrained optimization problems, which is the case of our problem. The method uses steps proportional to the opposite direction of the gradient of the objective function, at the current point. The adaptive step size is an acceleration technique, so that the convergence occurs faster. We define positive variables η_1 and η_2 as the steps that will be multiplied by $\frac{\partial f}{\partial \widehat{x}_1^{(k,j)}}$ and $\frac{\partial f}{\partial \widehat{x}_2^{(k,j)}}$ (analytically deduced) respectively. In our case, the



(a) Example of a world and the respective projected pattern.

(b) Projection of the pattern onto the camera's image.

Fig. 4: Example of a 3D scene used in the experimental results (a) and its respective camera's image (b). We use the same setup as Fig. 2.

adaptive component of the method increases η_1 (multiplying by $k_u > 1$) when the derivatives $\frac{\partial f}{\partial \widehat{x}_1^{(k,j)}}$ of the current and previous shearched points have the same sign, or decreases (multiplying by $0 < k_d < 1$) if they have opposite signs. The same reasoning is used for η_2, in respect to $\frac{\partial f}{\partial \widehat{x}_2^{(k,j)}}$.

Being an iterative method, at each iteration we will have a generic $\widehat{\mathbf{x}}^{(k,j)}$, which means that (20) will not be true. In addition to the objective function $f(.)$, we also use as stopping criteria $\|\mathbf{c}-\widetilde{\mathbf{c}}\|$, where we define $\widetilde{\mathbf{c}}$ as being the intersection of the ray that leaves $\widehat{\mathbf{x}}^{(k,j)}$ with the plane $z = c_3$, i.e.,

$$\widetilde{\mathbf{c}} = \widehat{\mathbf{x}}^{(k,j)} + \widehat{\beta}^{(k,j)} \widehat{\widetilde{\mathbf{d}}}_C^{(k,j)}, \;\; \widehat{\beta}^{(k,j)} = \frac{c_3 - \widehat{x}_3^{(k,j)}}{\widehat{\widetilde{d}}_{C,3}^{(k,j)}}. \qquad (21)$$

The reason why $\|\mathbf{c}-\widetilde{\mathbf{c}}\|$ is used as stopping criteria and not as objective function is because, despite having a physical interpretation simpler and more intuitive than our function, it's way more complex function than $f(.)$ and it would be computationally heavier to find the solution $\widehat{\mathbf{r}}_C^{(k,j)}$. The proposed algorithm is described in the appendix section.

After obtaining $\widehat{\mathbf{r}}_C^{(k,j)}$, as said before, one can obtain its projection $(\widehat{u}_C^{(k,j)}, \widehat{v}_C^{(k,j)})$ onto the camera's image using the camera's matrix $\mathbf{P}_C \in \mathbb{R}^{3\times 4}$, such that

$$\begin{bmatrix} \widehat{u}_C^{(k,j)} & \widehat{v}_C^{(k,j)} & 1 \end{bmatrix}^T \sim \mathbf{P}_C \begin{bmatrix} \widehat{\mathbf{r}}_C^{(k,j)} & 1 \end{bmatrix}^T. \qquad (22)$$

Doing this for all $\widehat{\mathbf{w}}^{(k,j)}$ we get the template image $\mathscr{T}(u,v)$, with the projection of all the discretized lines $\mathbf{l}^{(k)}$. An example of the template image is shown in Fig. 2(d) (for the projected pattern, projection lines, and discretized projection lines shown by Figs. 2(a)-(c)). The goal is to use this template image $\mathscr{T}(u,v)$ to perform the reconstruction of the world's scene, when an image $\mathscr{C}(u,v)$ is captured by the camera (through the mirror). This topic is covered in the next subsection.

C. Matching and reconstruction

To compute the 3D reconstruction, the goal is to use both the information of the projected pattern and an image acquired by the camera $\mathscr{C}(u,v)$. Since we have no knowledge about the world's scene, we will use the template image $\mathscr{T}(u,v)$, obtained in Sec. II-B, to estimate all the possible intersections of each line $\mathbf{l}^{(k)}$ with the scene, i.e., the position of each of $\widehat{\mathbf{w}}^{(k,j)}$. In Fig. 4(b), we show an example of a camera's image $\mathscr{C}(u,v)$ of the projected pattern (same pattern shown in Fig. 2). For this image, we simulate a scene as shown in Fig. 4(a).

With this information, in order to perform the matching, all we have to do is to find the closest point in $\mathscr{T}(u,v)$, to each point $(u_C^{(k)}, v_C^{(k)})$, in the camera's image using

$$\underset{n,m}{\text{argmin}} \quad \left(u_C^{(k)} - \widehat{u}_C^{(n,m)}\right)^2 + \left(v_C^{(k)} - \widehat{v}_C^{(n,m)}\right)^2 \qquad (23)$$

and, for the best solution n_* and m_* of (23), we reconstruct the 3D point associated with $(u_C^{(k)}, v_C^{(k)})$ as the corresponding 3D point $\widehat{\mathbf{w}}^{(n_*,m_*)}$.

In order to get better coverage of the world, without overloading the template image $\mathscr{T}(u,v)$ (which could cause some mismatches at (23)), one can use changing patterns which will create different incident rays in the mirror and, consequently, different $\mathbf{w}^{(k)}$ in the world. This can be done by capturing a sequence of $\mathscr{C}_i(u,v)$ (image frames), each of them corresponding to images captured by the camera, after the projection of the pattern to the scene world with different pattern template images $\mathscr{T}_i(u,v)$.

In conclusion, to reduce the noise in the reconstruction, we do an additional filtering to the 3D data. We consider the following procedure. For each of the k image points $(u_C^{(k)}, v_C^{(k)})$ (at the captured image $\mathscr{C}_i(u,v)$), we find the $N \geq 2$ closest 2D points in $\mathscr{C}_i(u,v)$ and consider its corresponding 3D reconstructed points and, with the $N+1$ resulting 3D points, we estimate a plane Π. This plane is computed using a RANSAC algorithm [20], in order to exclude outliers from the estimation of the plane. Then, the intersection point between the plane Π with the line $\mathbf{l}^{(k)}$ (projection line of $(u_C^{(k)}, v_C^{(k)})$) is computed and assigned to $\mathbf{w}^{(k)}$.

III. Experiments

To evaluate the proposed system, we need to simulate a world scene to project the pattern. An example of a world used in the tests is shown in Fig. 4(a). This scene is composed by four planes that form four different regions, allowing us to check how the reconstruction performs. As Ground-Truth (GT), we use the real $\mathbf{w}^{(k)}$ that can be computed for all k, as explained in Sec. II-A. In respect to the evaluation metrics, we used:

- The error in the reconstruction given by the euclidean distance between the reconstructed $\mathbf{w}^{(k)}$ and the respective GT points;
- The Field-Of-View (FOV) of the resulting system, according to the x-axis ($i = 1$) and to the y-axis ($i = 2$) direction. To calculate the FOV, for each of the directions, we use the origin and the two points with the largest value of $\frac{w_i^{(k)}}{-w_3^{(k)}}$ $(w_i^{(k)} > 0)$ and $\frac{-w_i^{(k)}}{-w_3^{(k)}}$ $(w_i^{(k)} < 0)$, where $i = 1,2$ (check Fig. 5).

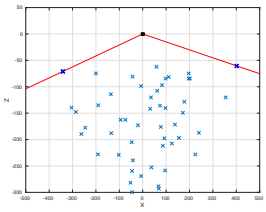
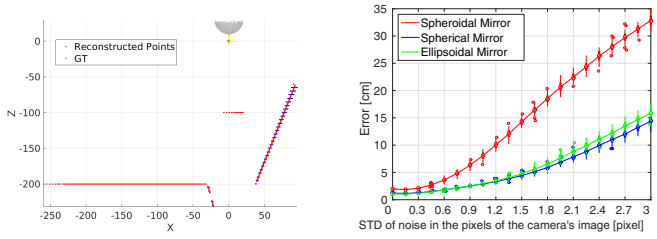


Fig. 5: Exemplification on how the FOV (x-axis direction) is calculated. The FOV considered is the smallest angle between the two red lines.



(a) (b)

Fig. 6: How the discretization and noise in the position of the pixels in $\mathscr{C}(u,v)$ affects the final results: (a) no discretization or noise in the process (when the reconstruction has no error only the GT points are visible); (b) the influence of discretization and noise in the 3D reconstruction.

In order to see how our approach performs, we designed several experiments to test its robustness, with several types of noise. Taking this in consideration, we choose 3 different mirror: a spherical one (with $A = B = 1$); a spheroid (with $A = B = 1.59$), and an ellipsoidal type mirror (with $A = 1.59$, $B = 1$). In all the mirrors, we choose C and D such that $C - \sqrt{D} > 0$.

For the evaluation, we choose four different tests. In the first one, we check how does the noise in the camera's image $\mathscr{C}(u,v)$ affects the reconstruction. In the remaining tests, we evaluate how the error and FOV are influenced by changes in the mirrors' D parameter; in the distance between the camera and the projector; and in the distance between the projector and the mirror. In all of the tests, the camera is pointed at $[0,0,C/2]^T$. To create the template points $\widehat{\mathbf{w}}^{(k,j)}$ (see (11)), we used a step $\alpha^{(k,j)}$, in order to consider points that verify $-300 < \widehat{w}_3^{(k,j)} < -10$ and steps of 5 in the $-z$ direction, for each $\mathbf{I}^{(k)}$.

A. The influence of noise in the camera's image

The goal of the first experiment is to test how the use of a real camera effects the final 3D reconstruction. Real cameras use discrete pixels. In addition, one has to take into account some noise associated with the pixels. Contrarily to the one used in the reconstruction shown in Fig. 6(a) (which show the reconstruction without noise). To simulate this situation, we add a normally distributed noise to $(u_C^{(k)}, v_C^{(k)})$, with mean 0 and standard deviation ranging from 0 to 3 pixels, for all k. The resulting new position of the pixels are then rounded to the nearest integer, in order to achieve the discretization.

In this simulation, the camera and the projector are at fixed positions $[0,-7.5,0]^T$ and $[0,0,0]^T$, respectively, and the three previous mentioned mirrors are used. The distance between the mirrors and the projector is 7.5, $D = 20^2$ and $C = 27.5$. Fig. 6(b) shows the respective results.

B. The influence of the mirror's D parameter

Changing the D parameter of the mirrors will change their size (radius in case of the spherical mirror) and, consequently, it will allow to reach points in different parts of world. In order to study this situation a test to evaluate how the reconstruction error and the FOV are affected was designed. In this test, the D parameter is increased from 10 to 30, with a step of 0.5, and the values of the two metrics are saved. The camera is located at $[0,-7.5,0]^T$, the projector at $[0,0,0]^T$, and distance between the mirrors and the projector is 7.5, which means that $C = 7.5 + \sqrt{D}$, for each D tested.

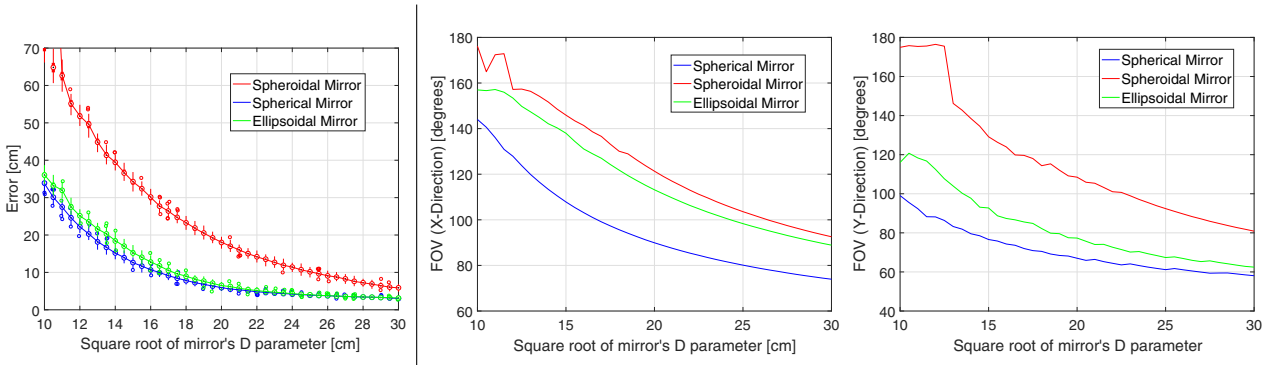
Discretization and normally distributed noise is introduced in the camera's image pixel, with null mean and standart deviation of 1.5 pixels. The results for the reconstruction error and the FOV with D is shown in Figs. 7(a) and 7(b) respectively.

C. The influence of the distance between the camera and the projector

The way the camera sees the world through the mirror changes as its distance to the projector is altered. This change will produce different results in the 3D reconstruction, since it dependes on the location of each point in both the camera and template's images. To observe the influence of this feature, we fixed the projector at $[0,0,0]^T$ and moved the camera from $[0,-2,0]^T$ to $[0,-20,0]^T$, with a steps of 0.5. For this purpose, we are considering $D = 20^2$ and $C = 27.5$. It is also considered a normally distributed noise, with null mean and standart deviation of 1.5, in the camera's image, as well as discretization. The results of the simulation are shown in Fig. 8(a), for the reconstruction error, and in Fig. 8(b), for the FOV.

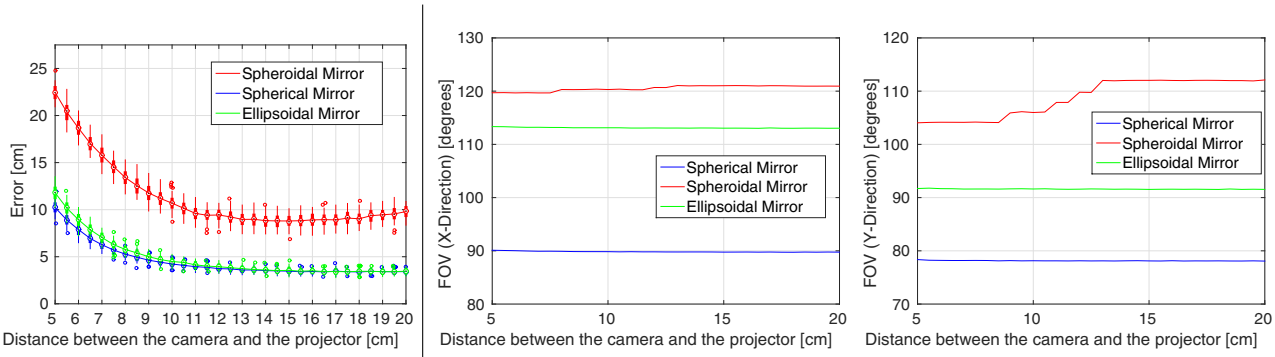
D. The influence of the distance between the mirror and the projector

Similarly to the case where the D parameter changes, when the distance between the mirror and the projector is modified, the reflections and points reached in the world will be different. In order to test this changes, the projector was put on a fixed location (at $[0,0,0]^T$), we set $D = 20^2$, and the distance between the mirror and the projector was increased from 5 to 17.5 (which implies changing C from 25 to 37.5), with a step of 0.5. The camera was located at $[0,-7.5,0]^T$ and we also took into account image discretization and a normally distributed noise with standard deviation of 1.5. Figs. 9(a) and 9(b) show the evaluation of the 3D reconstruction and FOV, respectively, when changing the respective change in the system's parameters.



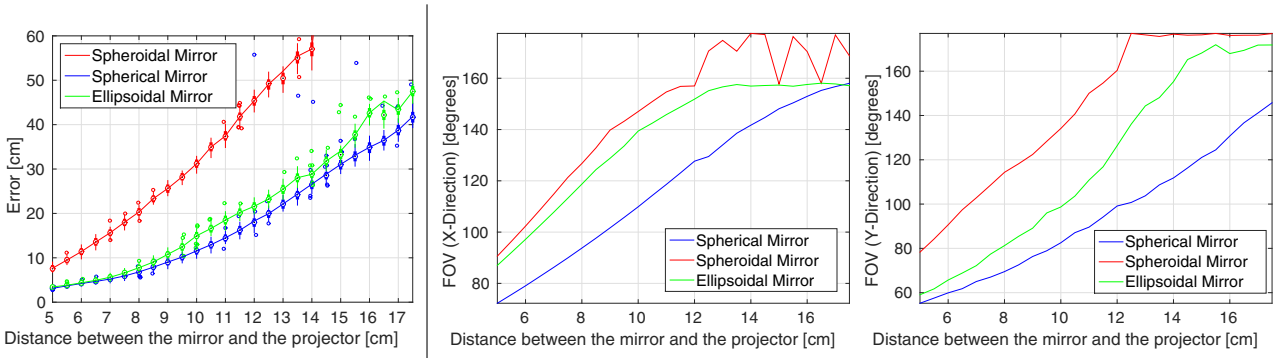
(a) (b)

Fig. 7: The influence of the mirror's D parameter (mirror's size – radius in case of the spherical mirror). in the 3D reconstruction (a) and FOV (b).



(a) (b)

Fig. 8: The influence of the distance between the camera and the projector in the 3D reconstruction (a) and FOV (b).



(a) (b)

Fig. 9: The influence of the distance between the mirror and the projector in the 3D reconstruction (a) and FOV (b).

IV. Conclusions

In this section we discuss the main results obtained from the experiments described in the previous section, and how they can influence our final goal of building an omnidirectional catadioptric RGB-D camera system, as well as what it's planned for future work.

A. Analysis of the results

From the four experiments realized, it was possible to take several important conclusions. The first experiment tested how a system with a real camera would behave, by considering discretization and noise pixel. We concluded that, adding the discretization of the pixels to the model created a baseline reconstruction error, that can be use as reference to the other experiments. Then, as expected, the error increased as the standard deviation of the noise in the pixels also did.

When considering the other experiments, one of the main conclusions obtained was that, in general, in order to increase the FOV (something that is essential in an omnidirectional system), one has to accept that the resulting reconstruction error will also increase. The only situation where this did

not happen was when the distance between the camera and the projector was increased, resulting in a decreasing reconstruction error, while the FOV remained approximately constant. This fact can be used in our favour, to lower the error but also without changing the FOV. However, one should take into account that this will enlarge the physical dimension of our system.

Other relevant situation was the fact that increasing the mirrors' D parameter resulted in a lower reconstruction error. However, this can be explained by the fact that, in these situations, the surface of the mirrors where reflections occur are approximately flat, turning the system more similar to a conventional RGB-D system, which results in a lower FOV.

To conclude, we can also see that the mirror parameters also affect both the errors in the reconstruction and FOV. From the three tested mirrors, the one that gave a more balanced performance was the ellipsoidal mirror. It obtained reconstruction errors almost as lower as the spherical mirror, while keeping wider FOV than the spherical and, in some cases, similar to the spheroidal mirror.

B. Final remarks and future work

In this work we addressed some problems related with building an omnidirectional catadioptric RGB-D camera, namely the problems associated with the estimation of the parameters associated with the depth channel. We develop a mathematical framework to solve this problem which, from the experimental result, allows as to say that such type of device in practice possible to develop.

In this paper we propose a novel omnidirectional RGB-D system, which uses a projector and cameras, both looking at the same quadric mirror. We also propose a template matching technique (including an approach for the projection of 3D points into the mirror's surface) for the reconstruction of the scene and, as result, for the computation of the depth channel of the sensor.

To conclude, as future work, we plan the construction of a prototype of the proposed system, that will allow us to put in practice the encouraging results obtained in the simulations.

REFERENCES

[1] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision," *Cambridge University Press*, 2004.

[2] H. Miles, S. Lee, O. Choi, and R. Horaud, "Time of Flight Cameras: Principles, Methods, and Applications," *Springer Briefs in Computer Science*, 2012.

[3] D. Fofi, T. Sliwa, and Y. Voisin, "A comparative survey on invisible structured light," *SPIE Electronic Imaging*, 2004.

[4] B. Freedman, A. Shpunt, M. Machline, and Y. Arieli, "Depth mapping using projected patterns," *US Patent 8,493,496*, 2013.

[5] B. Freedman, A. Shpunt, and Y. Arieli, "Distance-varying illumination and imaging techniques for depth mapping," *US Patent 8,761,495*, 2014.

[6] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, and W. Burgard, "An Evaluation of the RGB-D Slam System," *IEEE Proc. Int'l Conf. Robotics and Automation (ICRA)*, 2012.

[7] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A Benchmark for the Evaluation of RGB-D SLAM Systems," *IEEE/RSJ Proc. Int'l Conference on Intelligent Robot Systems (IROS)*, 2012.

[8] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "RGB-D mapping: Using Kinect-style depth cameras for dense 3d modeling of indoor environments," *Int'l Journal of Robotics Research*, 2012.

[9] K. Lai, L. Bo, X. Ren, and D. Fox, "A Large-Scale Hierarchical Multi-View RGB-D Object Dataset," *IEEE Proc. Int'l Conf. Robotics and Automation (ICRA)*, 2011.

[10] N. Silberman, P. K. D. Hoiem, and R. Fergus, "Indoor Segmentation and Support Inference from RGBD Images," *Proc. European Conf. Computer Vision (ECCV)*, 2013.

[11] R. Orghidan, J. Salvi, and E. M. Mouaddib, "Modelling and accuracy estimation of a new omnidirectional depth computation sensor," *Pattern Recognition Letters*, 2006.

[12] C. Zhang, J. Xu, N. Xi, Y. Jia, and W. Li, "Development of an omnidirectional 3d camera for robot navigation," *IEEE/ASME Int'l Conf. Advanced Intelligent Mechatronics*, 2012.

[13] D. Esparza, J. Barbosa, J. Ramos, and F. Rodriguez, "A Panoramic 3D Reconstruction System Based on the Projection of Patterns," *International Journal of Advanced Robotic Systems*, 2014.

[14] F. Zhou, B. Peng, Y. Cui, Y. Wang, and H. Tan, "A novel laser vision sensor for omnidirectional 3D measurement," *Optics & Laser Technology*, 2013.

[15] T. Jia, Y. Shi, Z. Zhou, and D. Chen, "3D depth information extraction with omni-directional camera," *Information Processing Letters*, 2015.

[16] R. Marani, V. Renò, M. Nitti, T. D'Orazio, and E. Stella, "A Compact 3D Omnidirectional Range Sensor of High Resolution for Robust Reconstruction of Environments," *Sensors*, 2015.

[17] Nuno Gonçalves, "On the reflection point where light reflects to a known destination on quadratic surfaces," *Optics Letters*, 2010.

[18] Amit Agrawal, Yuichi Taguchi, and Srikumar Ramalingam, "Beyond Alhazen's Problem: Analytical Projection Model for Non-Central Catadioptric Cameras with Quadric Mirrors," *IEEE Proc. Computer Vison and Pattern Recognition (CVPR)*, 2011.

[19] E. Fiesler and R. Beale, *Handbook of Neural Computation*. Oxford University Press, 1996.

[20] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Commun. ACM*, 1981.

APPENDIX

In this appendix we present the gradient method with adaptive step size for finding the reflection point $\widehat{\mathbf{r}}_C^{(k,j)}$.

Data: A, B, C, D, $\widehat{\mathbf{w}}^{(k,j)}$, $\mathbf{c}$
Choose initial $\widehat{\mathbf{x}}^{(k,j)}$ such that $\widehat{\beta}^{(kj)} > 0$;
Define values for k_u and k_d such that $k_u > 1$ and $0 < k_d < 1$;
Choose a *tolerance* for $\|\mathbf{c} - \widetilde{\mathbf{c}}\|$;
while $\varepsilon > tolerance$ **do**

$$\text{Compute } \frac{\partial f}{\partial \widehat{x}_1^{(k,j)}} \text{ and } \frac{\partial f}{\partial \widehat{x}_2^{(k,j)}};$$

$$\widehat{x}_i^{(k,j)} \leftarrow \widehat{x}_i^{(k,j)} - \eta_i \frac{\partial f}{\partial \widehat{x}_i^{(k,j)}}, \text{ for } i = 1,2;$$

$$\widehat{x}_3^{(k,j)} \leftarrow C - \sqrt{D - \left(A\widehat{x}_1^{(k,j)}\right)^2 - \left(B\widehat{x}_2^{(k,j)}\right)^2};$$

$$\widehat{\mathbf{d}}_C^{(k,j)} \leftarrow \widehat{\mathbf{x}}^{(k,j)} - \widehat{\mathbf{w}}^{(k,j)};$$

$$\widehat{\mathbf{n}}_C^{(k,j)} \leftarrow [2A\widehat{x}_1^{(k,j)}, 2B\widehat{x}_2^{(k,j)}, 2(\widehat{x}_3^{(k,j)} - C)]^T;$$

$$\widehat{\widehat{\mathbf{d}}}_C^{(k,j)} \leftarrow \widehat{\mathbf{d}}_C^{(k,j)} - 2\widehat{\mathbf{n}}_C^{(k,j)} \left(\frac{\widehat{\mathbf{n}}_C^{(k,j)} \cdot \widehat{\mathbf{d}}_C^{(k,j)}}{\|\widehat{\mathbf{n}}_C^{(k,j)}\|^2} \right);$$

$$\widehat{\beta}^{(k,j)} \leftarrow \frac{c_3 - \widehat{x}_3^{(k,j)}}{\widehat{\widehat{d}}_{C,3}^{(k,j)}};$$

$$\widetilde{\mathbf{c}} \leftarrow \widehat{\mathbf{x}}^{(k,j)} + \widehat{\beta}^{(k,j)} \widehat{\widehat{\mathbf{d}}}_C^{(k,j)};$$

$$\varepsilon \leftarrow \|\mathbf{c} - \widetilde{\mathbf{c}}\|;$$

if $d_i \frac{\partial f}{\partial \widehat{x}_i^{(k,j)}} > 0$, *for* $i = 1,2$ **then**

$$\eta_i \leftarrow k_u \eta_i, \text{ for } i = 1,2;$$

else

$$\eta_i \leftarrow k_d \eta_i, \text{ for } i = 1,2;$$

end

$$d_i \leftarrow \frac{\partial f}{\partial \widehat{x}_i^{(k,j)}}, \text{ for } i = 1,2;$$

end