Abstract

Inspection of an environment is often limited to a single video feed, provenient from a camera assembled on a mobile robot. This is known as the keyhole effect. In addition to this, the operator must often perform several tasks at the same time and specially in cluttered environments it can be hard to navigate the robot and at the same time keep the viewpoint of a camera aligned with a specific point of interest. This paper tackles this problem by proposing two mechanisms for assisting in teleoperation by positioning a pan-tilt camera. The goal consists in improving the teleoperator’s situation awareness. Field trials with a pan-tilt-zoom (PTZ) camera mounted on an all-terrain robot validate the proposed mechanisms.

1. Introduction

Getting mobile robots out of the semi-structured office and shop-floor environments has always been a goal of the robotics research community. In this context, substantial progress has been achieved in the last 15 years and with it, new interesting challenges are arising. Carlson and Murphy [3], go as far as to show that current state-of-the-art field robots are not able to operate without failures for more than 2 hours in the search & rescue domain. Recovery of such failures often requires human intervention. Apart from this, human judgement displays a critical element of a robot’s activity, specially in tasks that involve exploration, surveillance and reconnaissance. Thus, teleoperation is a requirement for mobile robots to successfully accomplish their tasks.

Cameras are a common denominator in most teleoperation interfaces (refer to [5] for a summary on methods for interfacing humans and robots highlighting this fact). The most common types of camera consist in spherical vision, fixed and pan-tilt cameras. Despite their ability to provide a panoramic view of the environment, spherical vision cameras are not efficient for teleoperation purposes as they can only offer low resolutions [17] and do not allow an environment to be analysed accurately. On the other hand, despite offering good resolution, fixed mounted cameras do not allow the operator to rapidly span its attention over the relevant aspects of a cluttered environment [18]. Pan-tilt cameras provide the operator with the ability of freely and accurately inspecting an environment, while simultaneously navigating the robot [12]. Hence, they are a better trade-off. Despite the advantages, several factors are important for a camera to be efficient. Deficient camera placement, narrow field of view, among others, can significantly harm the operator’s conception of the environment, resulting in cognitive mistakes and disorientation [8] (i.e. reduced situation awareness).

A teleoperation task normally engages the teleoperator in the subtasks of navigation and inspection [9]. Alternating between these two tasks, added to the previously mentioned issues, makes of robot teleoperation a very demanding task for the operator [8]. For example, in an uneven terrain, it requires the operator to constantly correct the camera’s position to maintain visual contact with a desired target. This results in distracting the operator from navigation, slowing down the task while inducing the operator in error.

The complexity of this problem drastically increases when data latency is high (e.g. as in space exploration rovers). This problem is typically solved by providing the robot with some autonomy that can aid teleoperation [13]. Shared viewpoint control [6], which consists in allowing an intermediate control of the viewpoint between the user and the rover, attentive navigation [19], coupled [7] and decoupled [9] camera controls, are state of the art hypotheses to assist on teleoperation. Besides the importance of the aforementioned techniques, the camera control systems (e.g. position/speed control) to support its manipulation also display an important role in a teleoperation process. For instance, latency (e.g. communications, processing, etc.) is an issue in a teleoperation task, where inspecting an environment through a camera can be quite confusing and it is therefore important to perform smooth movements with the goal of reducing image blur. It is therefore necessary to develop controllers which achieve a balance between motion smoothness and motion accuracy (e.g. ability to track a reference provided by the user).

Control systems are often crafted in order to cope with very specific and structured problems. The use of legacy
systems (e.g. commercial pan-tilt cameras) instead of custom made tools, is being more and more recurrent in robotics, requiring control systems that are capable of adapting to the limitations such systems offer. For instance, it is often impossible to attach a controller directly to a legacy system’s sensors/actuators which leads to the necessity of creating an interface to allow communication. This introduces limitations in the whole process (e.g. latency). Control systems must also be designed in order to cope with these cases.

The work presented in this paper consists in providing a pan-tilt camera with a control system to improve the operator’s experience by offering the possibility of sharing the control of the viewpoint. Therefore, two methods to help the operator to control the viewpoint of a camera are proposed in this paper: (i) pointing in azimuth and elevation and (ii) GPS pointing.

Section 2 presents an overview of the proposed controller. Section 3 adresses the problem of a position controller for a PTZ camera, followed by Section 4, which describes two tools for teleoperation aid on robot exploration. Section 5 describes the platform used to test the tools presented in this paper as well as experimental results which validate them. Finally, in section 6 conclusions are drawn and future work directions are given.

2 Controller Overview

Fig. 1 depicts the camera controller’s block diagram. Let us assume the following reference frames: World frame \{X_w, Y_w, Z_w\}, Site frame \{X_s, Y_s, Z_s\}, Rover frame \{X_rov, Y_rov, Z_rov\} and the Camera frame \{X_cam, Y_cam, Z_cam\}.

The world frame is fixed with regard to earth and its X-axis points North, Y-axis points west and Z-axis is parallel but opposite in sign to the local gravity vector. The site frame is the surface-relative frame of reference for all targeting and localisation operations. It has exactly the same orientation of the global reference frame, though, its origin is user defined. The rover frame is defined as follows: X-forward, Y-left and Z-up. The camera frame is identical to the rover frame, except that its origin is located in the camera’s Tilt axis (see Figure 2) (e.g. with regard to the rover frame, the camera frame is translated a \(T_z\) distance on the Z-axis).

Heading \(\psi_{rov}\), pitch \(\varphi_{rov}\) and roll \(\theta_{rov}\) represent the angle values relative to \(Z_{rov}, Y_{rov}\) and \(X_{rov}\), respectively, as depicted in Figure 2, and they are used to define the robot’s absolute attitude.

The rover’s attitude can be represented by a product of the three aforementioned rotations. The well known Tate-Bryant rotation sequence convention is used to represent the rover’s \(4 \times 4\) posture matrix, \(P_{rov}(\psi_{rov}, \varphi_{rov}, \theta_{rov})\), where the first rotation is applied around the Z-axis, the second around the Y-axis and finally around the X-axis,

\[
P_{rov}(\psi_{rov}, \varphi_{rov}, \theta_{rov}) = R_z(\psi_{rov})R_y(\varphi_{rov})R_x(\theta_{rov})
\]

where,

\[
R_z(\psi_{rov}) = \begin{bmatrix}
\cos(\psi_{rov}) & -\sin(\psi_{rov}) & 0 & 0 \\
\sin(\psi_{rov}) & \cos(\psi_{rov}) & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

\[
R_y(\varphi_{rov}) = \begin{bmatrix}
\cos(\varphi_{rov}) & 0 & \sin(\varphi_{rov}) & 0 \\
0 & 1 & 0 & 0 \\
-\sin(\varphi_{rov}) & 0 & \cos(\varphi_{rov}) & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

\[
R_x(\theta_{rov}) = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & \cos(\theta_{rov}) & -\sin(\theta_{rov}) & 0 \\
0 & \sin(\theta_{rov}) & \cos(\theta_{rov}) & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]
The kinematics module computes the reference pan and tilt angles which will define the input of the position controller, according to the user’s input. The inputs of this module are “mode” and the point of interest (POI), given in azimuth and elevation, or in GPS coordinates according to the mode in use, provided by the user, allowing the selection of pointing in azimuth/elevation or GPS pointing, respectively. In order to compute the correct angles to point at the desired objective, this module must also receive the posture and the localization of the robot, represented by matrix,

\[ \text{rov} = [x_{\text{rov}}, y_{\text{rov}}, z_{\text{rov}}, \varphi_{\text{rov}}, \theta_{\text{rov}}]^T \]

The position controller computes the error between the desired Pan and Tilt angles \( r_i(t) \) (where \( i = \text{P} \) if it is referred to the pan or \( i = \text{T} \) if it is the tilt axis) provided by the kinematics module and the position of the camera \( p_i(t) \) (an actual or a predicted value). The output of this block consists in the velocities of the pan and tilt actuators which will lead the camera to move into the desired position, \( r_i(t) \). This block as well as the state prediction block are detailed in Section 3.

### 3 Camera Pan/Tilt Position Control

As previously mentioned, inspecting an environment through a camera can be a daunting task. One way of reducing the effects of this problem consists in performing smooth transition between perspectives as it allows an operator to maintain situational context as the perspective changes [12]. This paper focuses on a controller to smoothly position a teleoperation camera.

A legacy surveillance PTZ camera was used. The impossibility of introducing the software controller into the camera, lead to the necessity of having the controller in a remote processor, requiring a communication interface that could be easily integrated between both. Several problems were identified: sluggish and error prone communication and excessive number of queries necessary to retrieve little information (e.g. pan/tilt positions).

Latency is a common problem in different types of applications (e.g. interplanetary communication, image processing, etc.). In this case, it is due to the fact that the controller is placed remotely from the camera and therefore required a communication interface constrained to the legacy camera. Both the identified problems with the camera consist in the introduction of delays in the control system. Neglecting such delays results in systems with slow responses as the gains need to be conservative to reach stability. The only way to improve performance at this level consists in recurring to state prediction to compensate time delays [11]. A necessary requirement to implement prediction is to have knowledge of the dynamic model of the system, in this case, the pan-tilt camera.

Ideally, at each cycle, the controller should have access to the pan and tilt angle values, which, in this case, results in an excessive number of queries. The predictive algorithm consists in estimating, either the pan or the tilt angles of the camera each cycle based on the period of the cycle and on the induced speed, either on the pan or on the tilt actuators. Then, in each cycle, one single query is made (thus obtaining alternately the Pan or the Tilt angles), being the second (and remainder) query replaced by a predicted value. As every two cycles, the information about the angle value of each axis is updated with actual sensorial values, errors do not accumulate over time.

By reducing the computation time it is possible to increase the gains of the controller without leading the system to undesired overshooting situations.

The core of the proposed controller consists of two Proportional-Derivative (PD) controllers, one per each camera actuator. The PD controller allows the obtaining of satisfactory results for a static reference, though, it proved to be insufficient for a dynamic one, which is important if the goal is to automatically adjust the camera’s viewpoint as the rover where it is mounted on moves. Therefore, an extra parcel was added to a conventional PD controller, where the variation of the reference is also taken into account, significantly improving the results in dynamic situations.

The output of the controller \( u_i(t) \) consists of the velocities to set the pan or tilt actuators in order to reduce the reference error, \( e_i(t) = r_i(t) - p_i(t) \), where \( p_i(t) \) refers to the position (true or the estimated) of axis \( i \) (see Fig. 3),

\[
\begin{align*}
  u_i(t) &= e_i(t)K_{ip} + \frac{de_i(t)}{dt}K_{id} + \frac{dr_i(t)}{dt}K_{ir} \quad (1)
\end{align*}
\]

where \( K_{ip}, K_{id} \) and \( K_{ir} \) are the proportional, the derivative and the varying reference gains, respectively.

A gain scheduling approach was used in order for the controller to cope with the non-linear behaviour of the camera [10]. The scheduling variables used to determine the operating regions which enable the different gains are the position errors between the camera’s actual position and the desired reference. Ideally, the camera should have fast movements when it is far of its desired position (i.e. high gains) and smoother movements when it

![Figure 3. Blocks diagram of the position controller](image-url)
is closer to the desired position (i.e. lower gains). Thus, three scheduling levels were established for the both camera actuators, a level for faster movements, another level for slower movements and an intermediate region with intermediate gains in order to guarantee smooth transitions in the movements of the camera. This approach allows the camera to present different behaviours properly parametrised for each type of situation.

4 Shared Viewpoint Control

Teleoperation experience can be improved significantly recurring to shared-control approaches [1, 4], which consists in mixing input provided by the teleoperator with information from the robot’s sensors. The goal is to provide the operator with mechanisms to define the desired orientation or a location to gaze at regardless of the robot’s posture and/or position. With this feature, it is possible to fixate on a desired object without effort for the operator.

In this context, two pointing mechanisms for aid on teleoperation are presented in this paper: (1) Pointing in Azimuth and Elevation and (2) Pointing to GPS point.

Azimuth and elevation are angles used to define the position of an object, relative to a reference plane. The azimuth corresponds to the compass bearing, relative to geographic north, of a point on the horizon. The elevation is measured between the horizon and the desired object, measured from the reference frame of the observer.

4.1 Pointing in Azimuth and Elevation (Az-El)

The goal of this mechanism consists in maintaining the camera aligned within a user defined azimuth and elevation. Obtaining the Pan and Tilt camera joints in order to aim at a desired Az-El can be seen as a reverse kinematics problem.

Let us use a unitary viewpoint vector represented on the earth frame by matrix \( \Omega \), to define the desired viewpoint,

\[
\Omega = \begin{bmatrix}
1 & 0 & 0 & \cos(-\eta)\cos(\varepsilon) \\
0 & 1 & 0 & \sin(-\eta)\cos(\varepsilon) \\
0 & 0 & 1 & \sin(\varepsilon)
\end{bmatrix}.
\]

where, \( \eta \) corresponds to the desired azimuth between the world frame and the camera’s line of sight and \( \varepsilon \) to the desired elevation between the world frame and the camera’s viewpoint vector.

The position of the pan and tilt actuators, matrix \( S \), can be determined from \( P_{rov} \) and \( \Omega \) as follows,

\[
S = \Omega P_{rov}
\]

This matrix represents the position of the pan and tilt camera actuators for a given Az-El. Matrix \( S \) is in the following form,

\[
S = \begin{bmatrix}
a_{11} & a_{12} & a_{13} & \alpha \\
a_{21} & a_{22} & a_{23} & \beta \\
a_{31} & a_{32} & a_{33} & \gamma \\
0 & 0 & 0 & 1
\end{bmatrix}.
\]

Before extracting the desired Pan and Tilt angles, it is necessary to transform matrix \( S \) into its row echelon form,

\[
S' = \begin{bmatrix}
1 & 0 & 0 & \alpha' \\
0 & 1 & 0 & \beta' \\
0 & 0 & 1 & \gamma'
\end{bmatrix}.
\]

The desired Pan, \( r_P \) and Tilt \( r_T \) angles are consequently given by,

\[
\begin{cases}
r_P(t) = \tan(\frac{\beta'}{\alpha'}) \\
r_T(t) = \sin(\gamma')
\end{cases}
\]

4.2 Pointing to GPS point

As previously mentioned, the goal of this mechanism consists in offering the user the possibility of gazing to a user defined GPS coordinate in the site frame. This mechanism can be particularly useful for surveillance purposes.

The first step of the GPS pointing mechanism consists in obtaining the camera’s desired Azimuth and Elevation between the robot’s position and the coordinates of the GPS point to inspect \( p = (x_p, y_p, z_p) \). Having the desired point identified, it is possible to extract its Az-El relative to the robot’s position as follows,

\[
\begin{cases}
\eta = \tan(\frac{x_{rov} - x_p}{y_{rov} - y_p}) \\
\varepsilon = \sin(\frac{z_p - z_{rov}}{\sqrt{(x_p - x_{rov})^2 + (y_p - y_{rov})^2}})
\end{cases}
\]

It is now necessary to determine the pan and tilt camera angles to maintain the just computed azimuth, \( \eta \), and elevation, \( \varepsilon \). For that purpose the method proposed in Section 4.1 is applied.

5 Prototyping

The following describes the test platform as well as the experiments carried out.

5.1 Test Platform Description

The test platform used to examine the implemented mechanisms is an all-terrain vehicle with four independently steered wheels, the Ares robot (see Fig. 4) [16]. The robot is equipped with a set of sensors for the following purposes,

- Camera Pan/Tilt proprioception: the used camera (a Pelco\(^1\) Esprit) is a custom PTZ surveillance camera whose base version already offers encoders for

\(^1\)http://www.pelco.com
measuring the attitude of its Pan and Tilt actuators. The Zoom is not considered for this work.

- **Attitude estimation**: for the purpose of assessing the robot’s posture, a magnetic compass with clinometers was used (a Honeywell HMR3000);

- **Localisation estimation**: Ares’ localisation estimation mechanism [14] depends on a wide variety of sensors. In short, optical encoders are used for a wheel odometer and a Videre Design STOC stereo camera for visual odometry [15]. These are used on a complementary way and are filtered with a DGPS system (a NovaTel OEMV-1) for global convergence.

5.2 Experimental results

As previously mentioned, simultaneously operating a robot and focusing a camera on a target can be a daunting task. Therefore, to analyze the performance of the proposed pointing mechanism an experiment in all-terrain was carried out. The test consisted in defining a GPS point for the camera to target while freely manoeuvring the Ares robot in an all terrain environment (see Fig. 5). This test allowed us to evaluate both the position controller as well as the performance of the GPS pointing mechanism.

The gains for each scheduling interval were tuned empirically and set as follows, for the pan controller,

$$\begin{cases} K_{TP} = 0.4, K_{TP} = 0.3, K_{TP} = 1 & \text{if } |e(t)| < 15^\circ \\ K_{TP} = 0.7, K_{TP} = 0.6, K_{TP} = 1 & \text{if } |e(t)| > 30^\circ \\ K_{TP} = 1, K_{TP} = 0.9, K_{TP} = 1 \end{cases}$$

and for the tilt controller,

$$\begin{cases} K_{Tc} = 0.5, K_{Tc} = 0.5, K_{Tc} = 1 & \text{if } |e(t)| < 15^\circ \\ K_{Tc} = 1.1, K_{Tc} = 0.8, K_{Tc} = 1 & \text{if } |e(t)| > 30^\circ \\ K_{Tc} = 1.5, K_{Tc} = 1.2, K_{Tc} = 1 \end{cases}$$

The results of the performance of the position controller are depicted in Figure 6 for the pan and in Figure 7 for the tilt actuator.

The average value for the error measured in the pan axis (see fig. 6(c)) consisted in a value of 5.19° with 4.98° of standard deviation. For the tilt axis (see fig. 7(c)) an average error of 2.40° with 3.20° of standard deviation was measured.

6 Conclusions and future work

The work presented in this paper is part of a wider project which aims at developing service robots [16], where the application of pointing mechanisms is often a requirement. A position controller was proposed as backbone to a technique for pointing a PTZ camera according to azimuth and elevation references, followed by a mechanism to track a GPS coordinate. Typically, to tackle this type of problem, custom made cameras and controllers are used. In this case, a legacy PTZ camera was selected and the presented mechanisms are specifically designed to cope with its limitations: legacy systems in general offer, in this case, sluggish and error prone communication. Despite the fact that the limitations offered by this camera do not allow the exploration of other interesting visual movements (e.g. saccadic movement), the proposed controller was capable of dealing with the limitations offered by the camera.

By taking out one of the queries, the employed state prediction method allowed for the reduction of the period of a cycle in approximately 25%, approximately 200 ms.
Figure 6. Results of the test for the position controller’s accuracy in all-terrain for the pan axis. 6(a) shows the reference for the pan actuator to follow, 6(b) displays the actual position of the pan actuator and 6(c) displays the error between the reference and the actual position of the pan actuator.

Figure 7. Results of the test for the position controller’s accuracy in all-terrain for the tilt axis. 7(a) shows the reference for the pan actuator to follow, 7(b) displays the actual position of the pan actuator and 7(c) displays the error between the reference and the actual position of the tilt actuator.
Despite having used a PTZ camera, the pointing mechanisms, as well as the developed controller, could be used to solve similar problems (e.g. antenna pointing [2]).

The two pointing mechanisms have been shown to provide significant help to the teleoperator, though, their impact in increasing situation awareness must be further studied. As future work, other methods for improving the control of a camera with the presented limitations can be tested, as trajectory prediction for example. Such pointing mechanisms can also be improved and tested with a camera with more degrees of freedom.

Acknowledgements

We thank Paulo Santos and Mário Salgueiro for their useful reviews and comments and to Vasco Santos for his support on preparing the experimental setup. This work was partially supported by FCT/MCTES grant No. SFRH/BD/27305/2006.

References


