Estimating world coordinates in perspective vision systems for humanoid robots

João Silva
joao.m.silva@ua.pt
Nuno Lau
nunolau@ua.pt
António J.R. Neves
an@ua.pt

Transverse Activity on Intelligent Robotics
IEETA/DETI University of Aveiro
3810-193 Aveiro

Abstract

The use of digital cameras on humanoid robots is very common. To extract information from camera images on these kind of platforms, we need to know the pose of the camera relative to our objects of interest. In the soccer application we project all the points of interest on the ground plane, in a 2D top view over the field of play. To accomplish this projection based on the pinhole camera model, we need to extract a set of values about the camera pose, based on the current position of the robot body and the camera position on the robot head. This document present a brief summary of the steps involved in such operation and the tool developed to support such steps.

1 Introduction

The use of cameras as main sensory information source for robots is nowadays almost a requisite, due to the flexibility of such source. To extract information from digital camera images, several methods of image processing are widely used. This results in the location of objects of interest on the image, which are pixel coordinates in a 2D representation of the world. In this document, we present the approach we use to extract information of point location on the world from the pixel coordinates provided by the vision analysis. The scenario used is robotic humanoid soccer on the RoboCup competition. The RoboCup is an international joint project to promote robotics and artificial intelligence. It includes several leagues, several of them related to soccer. Within these leagues, the Standard Platform League (SPL) is the one where this work is applied, a league where all the teams use the same robots, the humanoid NAO. These robots are equipped with 2 cameras which, in the version we use, do not have any overlap on their fields of view and cannot work simultaneously. Moreover, the robot controller is a Geode processor with a frequency of 500Mhz and possesses 256 MB of RAM. In the SPL, the objects of interest we use on the field are projected at the ground level. It is only thanks to this “restriction” that we are able to tackle the problem with a single camera without the need for more information.

2 Analysis of pinhole camera model

To accomplish the ground point projection, we use the pinhole camera model to analyze the geometrical relations of coordinates in 3D space and their projection into the camera 2D CCD. The first step was to make a static analysis of a vision system, in a way similar to the one presented for the CAMBADA team in [1]. Figure 1 presents two schematics of a camera system, side view and semi top view.

The several measurements involved are:

- h_{offset} - height of the camera relative to the ground;
- r_{offset} - radial distance from the camera to the robot center;
- α_{offset} - angular offset from the vertical axis to the camera center axis;
- x_{offset} - distance from the center of the robot to the point in the center of the image, projected on the ground;
- $ang_{\alpha n}$ - angle measured between α_{offset} axis and $pixel_n$;
- $ang_{\theta m}$ - angle measured between the robot frontal axis and $pixel_m$;
- $distance_{xn}$ - distance, measured on the ground, from the center of the robot to $pixel_n$ projected on the ground;

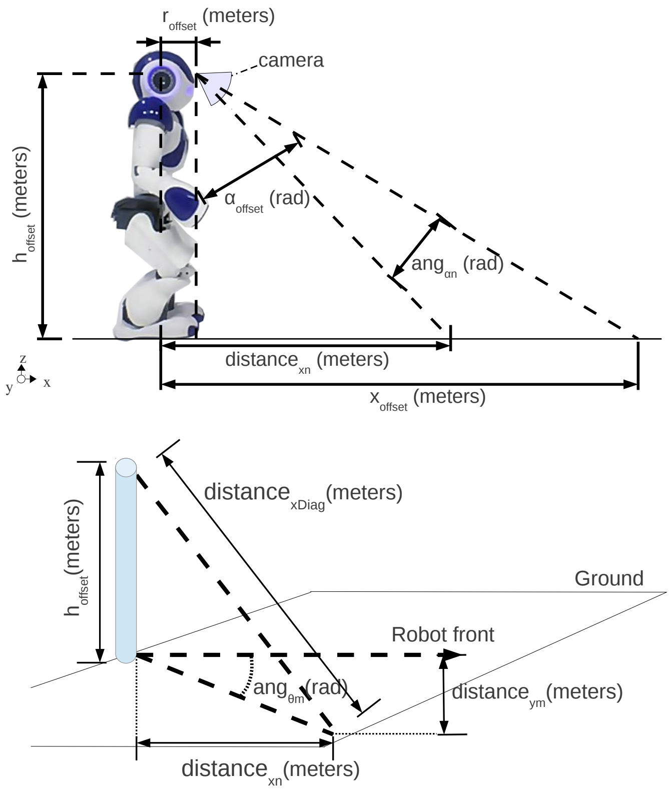

Figure 1: On top, side view schematic of the vision system. On bottom, a semi top view schematic of the vision system.

- $distance_{xDiag}$ - distance from the center of the camera to $pixel_n$ projected on the ground;
- $distance_{ym}$ - distance from the center of the robot to $pixel_m$ projected on the ground;
- $focal\ length$ - distance between lens and CCD;
- $pixel_n$ - number of pixels (n) along a CCD column;
- $pixel_m$ - number of pixels (m) along a CCD row;

Assuming the center of the CCD, we can easily verify that:

$$\alpha_{offset} = atan\left(\frac{x_{offset} - r_{offset}}{h_{offset}}\right) \quad (1)$$

From (1) we can generalize to an angle $ang_{\alpha n}$,

$$ang_{\alpha n} = \alpha_{offset} - atan\left(\frac{distance_{xn} - r_{offset}}{h_{offset}}\right) \quad (2)$$

Given the geometric properties of a camera vision system, we can relate an angle $ang_{\alpha n}$ and a pixel along a vertical column of the CCD,

$$pixel_n = \frac{tan(ang_{\alpha n}) \times focal\ length}{pixel\ height} \quad (3)$$

From (2) and (3), we obtain

$$pixel_n = tan\left[\alpha_{offset} - atan\left(\frac{distance_{xn} - r_{offset}}{h_{offset}}\right)\right] \times \frac{focal\ length}{pixel\ height} \quad (4)$$

With some manipulation of 4, the distance corresponding to each pixel can be found by

$$distance_{xn} = r_{offset} + h_{offset} \times tan\left[\alpha_{offset} - atan\left(\frac{pixel_n \times pixel\ height}{focal\ length}\right)\right] \quad (5)$$

We can thus obtain the distance, from the robot center, of any point on the image, using (5).

Following a similar analysis and based on the schematic of Fig. 1, we know relation between $ang_{\theta m}$ and distances $distance_{xn}$ and $distance_{ym}$ projected on the ground:

$$ang_{\theta m} = atan\left(\frac{distance_{ym}}{distance_{xn} - r_{offset}}\right) \quad (6)$$

Since we have a height associated, and the horizontal angle has a relation with both the distances (XX and YY), we can derive the following

$$distance_{ym} = pixel_m \times distance_{xDiag} \times \frac{pixel\ size}{focal\ length} \times cos\left[\alpha - atan\left(\frac{distance_{xn} - r_{offset}}{h_{offset}}\right)\right] \quad (7)$$

3 NAO vision system

The initial analysis was based on a system with a fixed camera, with its vertical axis aligned with the robot front and horizontal axis parallel to the ground. In the case at hand, the NAO robot, the camera is placed on the head, which is a mobile part of the robot. Thus, in this case, we have to consider that the camera have rotation on all three axis XX, YY and ZZ, commonly known as roll, pitch and yaw respectively.

Looking at a camera in space, and considering XX the camera focal axis with the YY/ZZ plane on the CCD, we can see that if we apply a roll angle to the camera, we have a rotation of the pixels. The first step of the coordinates estimation is to rotate the given pixel by the roll angle of the camera to correct that distortion.

The pitch analysis basically results in the α_{offset} that allows us to get the coordinates relative to the current image. Through expressions (5) and (7) we estimate the coordinates of the given pixel cosidering that the front of the robot is the current direction of the camera.

Finally, knowing the yaw of the camera, we can rotate the given point, already in ground plane coordinates, and obtain the final coordinates, relative to the robot front.

3.1 Extracting camera angles

To get the necessary angles of the camera on the space, we make use of the robot kinematics, combined with the use of the inertial unit present on the robot. Moreover, each camera of each robot has different angles relative to the head, since the construction of the robot does not guarantee precise values. Thus, we need to check the cameras roll, pitch and yaw angles on the head.

To help in extracting this information and to visually confirm both these parameters and the results of the estimations presented, a visual tool was created. This tool has a set of parameters necessary to calculate the ground coordinates of the image pixels and can run with a live feed from the robot camera, as well as a live feed from the robot sensors which allows us to verify also the kinematic model estimations, in an offline, static scenario. For the initial stage of the tool, we solved (5) and (7) in relation to $pixel_n$ and $pixel_m$, so that we can project a grid with known distances on the image pixels according to our model. Figure 2 presents some screen shots of the application with such a grid. In the example, the squares have a 36mm side and the bottom line is at 311mm from the robot center, manually measured with tape. The robot is placed in such a way that it is parallel to the chessboard.

4 Results

The developed tool was a need that came up when starting to extract the pixel ground projection information. The main need was initially to help the development and eventual corrections of the expressions, based on the knowledge of what the grid should be like over the image of a known

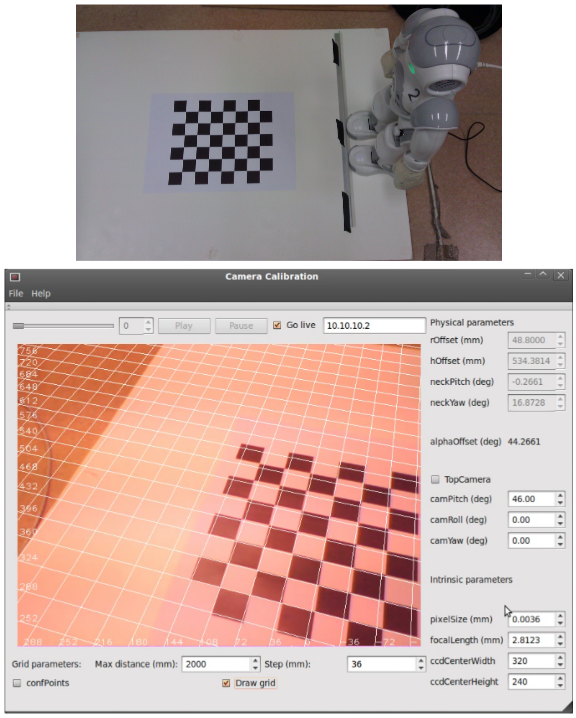

Figure 2: On top, the setup with known measures. On bottom, the 36mm grid projected on the image, with some pitch and yaw applied to the robot head.

environment. It is also a tool to try and test the angular parameters of the camera, also allowing to test the intrinsic parameters focal length, pixel size and CCD center.

Figure 3 illustrates the projection of points over the center circle of the soccer field when applying the expressions over the theoretical camera placement on the robot head and with the correction estimated through the camera calibration tool.

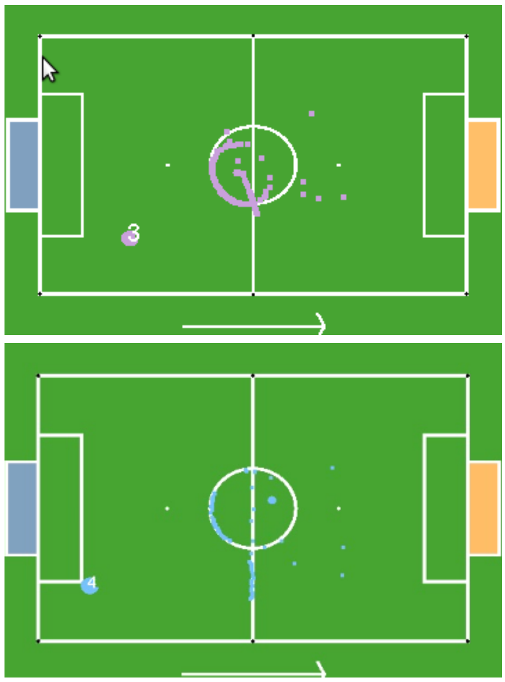

Figure 3: On top, the projection of points over the center circle using the projection without camera angle corrections. On bottom, the projection of a similar set of points now using the correction of the camera angles extracted through camera calibration tool.

References

[1] António J. R. Neves, Daniel A. Martins, and Armando J. Pinho. Obtaining the distance map for perspective vision systems. In *Proc. of the ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing*, Porto, Portugal, October 2009.