Robotics and Autonomous Systems 61 (2013) 1615–1625

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

3D monocular robotic ball catching

Vincenzo Lippiello *, Fabio Ruggiero, Bruno Siciliano

PRISMA Lab, Dipartimento di Ingegneria Elettrica e Tecnologie dell'Informazione, Università degli Studi di Napoli Federico II, via Claudio 21, 80125, Napoli, Italy

HIGHLIGHTS

- The robotic 3D ball catching problem is solved by using a monocular visual system.
- Comparison with the ground-truth provided by an OptiTrack motion-capture system.
- An industrial robot manipulator endowed of a camera mounted inside the gripper.
- Continuous refinement of the interception point with nonlinear estimation algorithm.
- Initial starting solution is provided by a fast linear estimating process.

ARTICLE INFO

Article history:
Received 7 February 2013
Received in revised form
3 June 2013
Accepted 21 June 2013
Available online 11 July 2013

Keywords:
Ball catching
Monocular camera
Visual servoing
Trajectory estimating

ABSTRACT

A new method to catch a thrown ball with a robot endowed with an eye-in-hand monocular visual system integrated into a gripper is proposed. As soon as the thrown ball is recognized by the visual system, the camera carried by the robot end-effector is forced to follow a baseline in the space so as to acquire an initial dataset of visual measurements from several points of view, providing a first estimate of the catching point through a linear estimation algorithm. Hereafter, additional measurements are acquired to constantly refine the previous estimate by exploiting a nonlinear estimation algorithm. During the robot trajectory, the translational components of the camera are controlled in such a way as to follow the planned path to intercept the ball, while the rotational components are forced to keep the ball into the field of view. Experimental results performed on a common industrial robotic system prove the effectiveness of the presented solution.

1. Introduction

Advanced robotic systems, which are required to perform quick reactions in response to visually perceived movements in a partially structured environment, are no doubt a good benchmark for testing of new control algorithms and new estimating/predicting processes. The challenging scenario of the robotic ball catching problem has been extensively considered in the literature for experimental testing of the above capabilities.

The state of the art presented in Section 2 shows that most of the existing systems use either a stereo visual configuration to solve the 3D ball catching problem or a single camera for the 2D sub-problem. In the former case, the 3D tracking of the ball takes advantage from having the possibility to exploit the epipolar constraint in the two available images; in the latter case, only 2D information can be directly retrieved from the single available image. Nevertheless, in both situations, in order to obtain a successful catch, high frame rate and accurate cameras are required to have a fast and precise estimate of the ball trajectory. By employing just one camera it is possible to reduce the total cost of the set-up. Moreover, the computational cost to elaborate at high frame rate one image is lower than elaborating at the same frequency two images. Hence, when high frame rate is a strict constraint, using only one camera also saves computational resources. However, some improvements in the controller and in the prediction algorithm should be introduced to solve the problem in 3D.

In this paper, the robotic 3D ball catching problem is solved by using a monocular visual system. A standard industrial robot manipulator is equipped with a CCD camera mounted directly on the manipulator end-effector. The proposed control law is composed of a continuous refinement of the ball interception point through a nonlinear estimation algorithm, whose initial starting condition is provided by a fast linear estimating process. An initial camera motion is thus commanded along a suitable baseline so as to collect a sufficient initial number of visual data from different points of view and provide such initial estimate. Experimental results demonstrate the effectiveness of the presented solution. The present work extends what already presented by the authors

* Corresponding author. Tel.: +39 0817683635.
E-mail addresses: vincenzo.lippiello@unina.it (V. Lippiello), fabio.ruggiero@unina.it (F. Ruggiero), bruno.siciliano@unina.it (B. Siciliano).

0921-8890/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.robot.2013.06.008

in [1]. With respect to the past work, in this paper the analysis of the employed controller is more detailed and the stability proof is provided. Moreover, the estimating process has been improved by introducing a recursive outliers rejection algorithm that improves the measurement dataset employed for the estimation process. Finally, the policy employed for the interception point selection has been detailed considering the physical limits of the robot, while the performance of the proposed trajectory estimator is shown through a comparison with the ground-truth provided by an OptiTrack motion-capture system.

The outline of the paper is as follows: after the presentation of related work, an overview of the proposed algorithm is provided and the image processing is briefly revised. Then, the partitioned visual-servoing controller adopted during the catching motion is presented along with the stability proof. Further, the linear and nonlinear estimation algorithms are formalized. Finally, the adopted set-up, the achieved experimental results and the comparison with the available ground-truth are shown and critically discussed.

2. Related work

Several research works address problems related to the catch of thrown objects and to the estimate of their trajectories. The approaches described in this section can be categorized as follows: stereo visual systems that deal with ball-catching tasks; monocular visual systems performing catching operations in a plane; monocular visual systems that estimate the object trajectory and perform catching operations in the 3D space; estimators dealing with Chapman hypothesis; systems that take into account the forces exchanged between the gripper and the thrown object; dynamic non-prehensile catching tasks; neural networks; virtual reality applications.

2.1. Stereo visual systems

Vision systems employing two (or more) cameras benefit by using the triangulation method to reconstruct the 3-D position of the ball [2–4]. However this requires an accurate calibration procedure and a sophisticated elaboration hardware.

A stereo vision system combined with an observer with a variable strength filter and an error estimator are employed in [5] to track and catch a thrown ball. An initial motion algorithm is chosen to maximize the response time so as to begin the motion of the arm as soon as the first visual data is taken.

A stereo visual system, an extended Kalman filter and a prediction algorithm are employed in [6] to build a robotic ball catcher. Without using specialized hardware, only using off-the-shelf components, the authors employ a stereo visual system to track a fast flying object and to catch it with a net mounted upon a robotic arm. In order to detect the ball, the difference between the actual image and some reference images is computed using a threshold method. Lately, a mobile humanoid and a gradient method to detect circles in the images are employed in [7].

An inexpensive and uncalibrated camera is exploited in [8] to track a rolling ball before it falls from a table. A robotic arm is programmed to catch it through an attractor-based dynamics that autonomously generates temporally discrete movements and sequences for the robot end-effector.

Only 2D visual information given by a stereo vision system is employed in [9] to achieve position control of a 3D robotic arm and catch a thrown object. The control is applied to achieve simultaneously 2D tasks defined directly on the image planes of the cameras: the 3D task is considered as accomplished if all the 2D tasks are simultaneously fulfilled. However, such working condition cannot guarantee the catch, since there is no estimate of the 3D ball motion. Moreover, no prior knowledge about the configuration and dimension of the robotic arm is needed as well as no information about camera parameters is requested (i.e., the robot Jacobian and camera calibration are estimated on-line).

A planar robotic arm, a stereo visual system and a DSP equipment are utilized in [10] to predict the right falling place of moving balls through a Lagrangian interpolation formula.

2.2. Monocular visual systems: catching in a plane

Monocular visual systems have easier calibration procedures, but more effort has to be put in the 3D reconstruction of the scene.

A camera and a prediction-based control system are employed in [11] to catch a mouse moving on a plane. A single camera is employed to localize the moving part, but since that is free to change the velocity and the acceleration of its motion in the whole plane, then a continuous re-planning of the path of the manipulator is required. The catch is always performed along a predetermined catching line, and for this reason the target object should cross such a line in a finite amount of time.

An experiment consisting in the catch of a ball moving on a table with a robot manipulator is carried out in [12]. Two distinct visual servoing architectures are implemented: position-based and image-based visual servoing. The catch is always performed along a straight line on the table and thus the precise catching point is determined with the intersection between such a line and the predicted trajectory of the ball, which is observed by a camera mounted on the robot end-effector.

2.3. Monocular visual systems: catching in 3D

The 3D position and velocity of a thrown projectile are estimated in [13] through the analysis of a sequence of images taken by a single camera. A least-squares algorithm is employed to determine the state of such projectiles from their apparent trajectories and considering a model of the motion without air drag.

A recursive least-squares algorithm is even used in [14] to estimate the trajectory of a ball with an eye-to-hand visual system. Catching is performed through a combination of image-based and position-based visual servoing.

This case best fits the work here presented. With respect to the cited works, a monocular eye-in-hand configuration is considered in this paper, hence the camera is mounted directly on the end-effector of the robot. In this way, the control is in charge of maintaining the ball inside the camera field of view, which instead is not possible in a eye-to-hand configuration. However, the visual system has to cope with the change of resolution of the ball in the image during the throw and with possible blur motions. Moreover, with respect to the previous works, a more realistic model of the ball motion is here employed, including also the air drag factor, along with a more sophisticated nonlinear estimation process.

2.4. Estimators dealing with Chapman hypothesis

The Chapman strategy to catch a ball is introduced in [15], where it is stated that a fielder should run at a proper speed to maintain a constant increasing rate of the tangent of the ball elevation angle.

Reinforcement learning models are exploited in [16] to better understand the perceptual features that guide a fielder to learn how to catch a flying ball. For this reason, the authors implemented a system which learns both how to keep constant the increasing rate of the tangent of the elevation angle and how to use the velocity of the ball perpendicular to the fielder to decide whether to run forward or backward.